



Faculty of Engineering Science
Department of Mechanical Engineering
Centre for Industrial Management,
Traffic and Infrastructure

Optimal Resource allocation and Budgeting in Libraries

Lorena Catalina SIGUENZA-GUZMAN

Dissertation presented in partial fulfillment of the requirements
for the degree of Doctor in Engineering

Supervisors:

Prof. Dr. Ir. Dirk Cattrysse
Prof. Dr. Henri Verhaaren

Chairman:

Prof. Dr. Ir. Hugo Hens

Examination Committee:

Prof. Dr. Alexandra Van den Abbeele
Prof. Dr. Ir. Joos Vandewalle
Prof. Dr. Raf Dekeyser
Prof. Dr. Ir. Juan Pablo Carvallo

August 2015

© 2015 KU Leuven, Science, Engineering & Technology

Uitgegeven in eigen beheer, Lorena Siguenza Guzman, Leuven.

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaandelijke schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm, electronic or any other means without written permission from the publisher.

*No university in the world has ever risen to greatness without a
correspondingly great library...*

When this is no longer true, then will our civilization have come to an end.

(Lawrence Clark Poell)

Acknowledgment

The completion of a PhD thesis is never an individual effort and I would like to express my sincere gratitude to all people and institutions that made this dissertation possible. I am quite convinced that this space is not enough to thank everyone who has helped me, not only in the academic and research aspects but has also been present in the personal matters.

First and foremost, I would like to thank God for all his blessings and love. I am convinced that he has never left me through this journey.

I would like to acknowledge and express my sincere gratitude to my supervisor Prof. Dirk Cattrysse for his guidance since my first attempts at formal research. I have received his support, supervision and guidance on both research and personal aspects. I would like to thank him for encouraging my research and for allowing me to grow as a research scientist.

Besides my supervisor, I would like to thank the rest of my Examination Committee: Prof. Rik Verhaaren, Prof. Alexandra Van den Abbeele, Prof. Joos Vandewalle, Prof. Raf Dekeyser and Prof. Juan Pablo Carvallo for serving as my committee members. Rik, thank you very much for adding a different perspective to my Engineering background. Thank you Alexandra for your advice, support, comments and “cheering”. Prof. Vandewalle, thank you for your patience and willingness to help me, especially in the last chapters. Raf, thank you for agreeing to be part of my doctoral committee, but mainly for sharing your experience during all our periodical VLIR meetings. Juan Pablo, not only thank you for being part of my Doctoral Committee but also for your constructive comments which have improved both my text and my insights on the topic.

Likewise, I would like to express my deepest gratitude to the University of Cuenca, with special reference to its authorities: former Rector Jaime Astudillo, current Rector Fabian Carrasco and current Vice-rector Silvana Larriva, who support me since the very beginning. I am deeply thankful to the Institutional University Cooperation program from the Flemish Interuniversity Council (VLIR-IUC) on behalf of their authorities, Guido Wyseure, Fabian Leon, Miguel Cordero, Tupac Calfat and Piet Wostyn, and the National Secretariat of Higher Education, Science, Technology and Innovation of Ecuador (SENESCYT) for providing the financial support to complete this research.

I wish to thank the libraries that were part of this research and all the people who provided me with the information, data and facilities required during the data collection period.

I would like to express my gratitude to my former and current colleagues of the Center for Industrial Management (CIB) for their professionalism help and friendship. Special thanks to Pablo, Francesco, Jasmine, Reginald, Thijs, Ali D, Ali S, Corrinne, Nagham, Tugba and Mathias, my office mates with whom I shared the joys and frustrations of technical and not-so-technical issues. I

would like to also thank Jef, Wei, Xin, Marco, Paola, Farzad, Evy, Eva; Pieter, Karel and Peter, from whom I also keep very nice memories. I cannot finish this part without mentioning my new “cuencanos” colleagues: Andres, Elina and Ivan, who came to the department to reinforce the Ecuadorian touch; thank you for re-introducing me to the Ecuadorian world.

I want to thank to the Belgian team: Raf, Ludo and Frederic for our fruitful discussions and advice to improve the library of the University of Cuenca. I wish to thank the KU Leuven Arenberg Campus Library staff, especially to Hilde Van Kiel, Christophe Nassen, Ria Vanhove, Marc Verbrugge, Edwin Smellinckx, and Tim Therry for their big support. I would like to extend my sincerest thanks and appreciation to the Ecuadorian team: Rocio, Victor, Mauricio, Andres, Paul, Carlos, Valeria and Wilson, without your assistance this research could not have been proved and implemented in Cuenca.

I also thank my friends (too many to list here but you know who you are!) for providing the support and friendship that I needed. Special acknowledgments to the Cuencanos community, Ecuadorian community, and Cuban community for sharing joyful moments with me and my family. To my family in Leuven: Adri and Loli. You prove me that friends are the family we choose.

I especially thank my mom, dad, and sister who have always been my pillars of strength. My parents who constantly provide me unconditional love and care. I love them so much, and I would not have made it this far without them. My sister has been my endless source of happiness, the proof that God knows what is best for us. My family, including my in-laws, who have always supported me throughout this journey, thank you.

During these past years, I have been sharing the good and bad moments with my best friend, soul-mate, and husband. I married the best person out there for me. There are no words to express how much I love you. You have been a true and great supporter that has unconditionally loved me during my good and bad times. You have had faith in me even when I felt like digging a hole and crawling into one because I did not have faith in myself. These past years have not been easy, both academically and personally, I truly thank you for sticking by my side. The best outcome from this life experience has been my beloved son Martin and our new and little pretty gift, who made me live the experience of motherhood, and provided me a pleasurable source of distraction. You both are my shelter, my truth.

Abstract

Libraries since their inception 4000 years ago have been in a process of constant change. Although, changes were in slow motion for centuries, in the last decades, academic libraries have been continuously striving to adapt their services to the ever-changing user needs of students and academic staff. In addition, e-content revolution, technological advances, and ever-shrinking budgets have obliged libraries to efficiently allocate their limited resources among collection and services. Unfortunately, this resource allocation is a complex process due to the diversity of data sources and formats requiring analysis prior to decision-making, as well as the lack of efficient integration methods.

The main purpose of this study is to develop an integrated model that supports libraries in making optimal budgeting and resource allocation decisions among their services and collection by means of a holistic analysis. To this end, a combination of several methodologies and structured approaches is conducted. Firstly, a holistic structure and the required toolset to holistically assess academic libraries are proposed to collect and organize the data from an economic point of view. A four-pronged theoretical framework is used in which the library system and collection are analyzed from the perspective of users and internal stakeholders. The first quadrant corresponds to the internal perspective of the library system that is to analyze the library performance, and costs incurred and resources consumed by library services. The second quadrant evaluates the external perspective of the library system; user's perception about services quality is judged in this quadrant. The third quadrant analyses the external perspective of the library collection that is to evaluate the impact of the current library collection on its users. Eventually, the fourth quadrant evaluates the internal perspective of the library collection; the usage patterns followed to manipulate the library collection are analyzed.

With a complete framework for data collection, these data coming from multiple sources and therefore with different formats, need to be integrated and stored in an adequate scheme for decision support. A data warehousing approach is secondly designed and implemented to integrate, process, and store the holistic-based collected data. Ultimately, strategic data stored in the data warehouse are analyzed and utilized for different purposes including the following: 1) Data visualization and reporting is proposed to allow library managers to publish library indicators in a simple and quick manner by using online reporting tools. 2) Sophisticated data analysis is recommended through the use of data mining tools; three data mining techniques are examined in this research study: regression, clustering and classification. These data mining techniques have been applied to the case study in the following manner: predicting the future investment in library development; finding clusters of users that share common interests and similar profiles, but belong to different faculties; and predicting library factors that affect student academic performance by analyzing possible correlations of library usage and academic performance. 3) Input for optimization models, early experiences of developing an optimal resource allocation model to

distribute resources among the different processes of a library system are documented in this study. Specifically, the problem of allocating funds for digital collection among divisions of an academic library is addressed. An optimization model for the problem is defined with the objective of maximizing the usage of the digital collection over-all library divisions subject to a single collection budget. By proposing this holistic approach, the research study contributes to knowledge by providing an integrated solution to assist library managers to make economic decisions based on an “as realistic as possible” perspective of the library situation.

Korte Inhoud

Sinds hun oprichting 4000 jaar geleden, zijn bibliotheken voortdurend in een proces van verandering geweest. Hoewel de veranderingen eeuwenlang uiterst traag gebeurden, moeten wetenschappelijke bibliotheken in de laatste decennia zich voortdurend inspannen om hun diensten aan te passen aan de steeds veranderende gebruikersbehoeften van de studenten en het wetenschappelijk personeel. Daarnaast zijn bibliotheken door de e-content revolutie, de technologische vooruitgang, en de steeds krimpende budgetten verplicht om hun beperkte middelen efficiënt te verdelen tussen investeringen in de collectie enerzijds en dienstverlening anderzijds. Helaas is deze toewijzing van middelen een complex proces. Oorzaak hiervan is de diversiteit van de gegevensbronnen en formaten die moeten worden geanalyseerd voorafgaand aan de besluitvorming, alsmede het gebrek aan efficiënte integratiemethoden.

Het belangrijkste doel van deze studie is om een geïntegreerd model te ontwikkelen dat bibliotheken ondersteunt in het maken van een optimale begroting en de beslissingen over de toewijzing van middelen onder hun diensten en de collectie door middel van een holistische analyse. Hiertoe wordt een combinatie van verschillende methodes en gestructureerde benaderingen uitgevoerd. Als eerste worden een holistische structuur en de vereiste toolset om academische bibliotheken holistisch te kunnen beoordelen, gepresenteerd voor het verzamelen en ordenen van de gegevens vanuit economisch oogpunt. Hierbij wordt gebruik gemaakt van een vierledig theoretisch kader, waarin het bibliotheeksysteem en de collectie worden geanalyseerd vanuit het perspectief van de gebruikers en interne stakeholders. Het eerste kwadrant correspondeert met het inwendige perspectief van het bibliotheeksysteem dat de prestatie van de bibliotheek en de kosten en middelen verbruikt door bibliotheekdiensten analyseert. Het tweede kwadrant evalueert het externe perspectief van het bibliotheeksysteem; de perceptie van de gebruiker over de kwaliteit van diensten wordt beoordeeld in dit kwadrant. Het derde kwadrant analyseert het externe perspectief van de bibliotheekcollectie om het effect van de huidige bibliotheekcollectie op de gebruikers te evalueren. Tenslotte evalueert het vierde kwadrant het intern perspectief van de collectie van de bibliotheek; de gebruikspatronen die worden gevolgd om de bibliotheekcollectie te veranderen, worden geanalyseerd.

Met een volledige kader voor de gegevensverzameling moeten deze gegevens die komen uit verscheidene bronnen en derhalve met verschillende formaten, worden gecombineerd en verwerkt tot een geschikt beslissingsondersteunend schema. In een tweede stap wordt een data-warehousingaanpak ontworpen en geïmplementeerd om de op holistische basis verzamelde data te integreren, verwerken en op te slaan. Tenslotte worden de strategische gegevens die opgeslagen zijn in het data-warehouse geanalyseerd en gebruikt voor verschillende doeleinden, waaronder de volgende: 1) Data-visualisatie en -rapportering worden gepresenteerd om bibliotheekmanagers de mogelijkheid te geven bibliotheekindicatoren te publiceren op een eenvoudige en snelle manier met behulp van online rapporteringssystemen. 2) Geavanceerde gegevensanalyse wordt uitgevoerd door middel van datamining; drie dataminingtechnieken worden beschouwd in dit onderzoek: regressie, clustering en classificatie. Deze dataminingtechnieken worden op de volgende manier

toegepast op een case study: het voorspellen van de toekomstige investeringen in de ontwikkeling van de bibliotheek; het vinden van clusters van gebruikers die gemeenschappelijke belangen en soortgelijke profielen hebben, maar behoren tot verschillende faculteiten; en het voorspellen van elementen uit de bibliotheek die academische prestaties van studenten beïnvloeden door het onderzoeken van mogelijke correlaties van de bibliotheekgebruik en academische prestaties. 3) Input voor optimaliseringsmodellen, de eerste ervaringen met het ontwikkelen van een optimaal brontoewijzingsmodel om middelen te verdelen over de verschillende processen van een bibliotheekstelsel worden beschreven in deze studie. Concreet wordt het probleem van de verdeling van de middelen voor de digitale collectie onder de afdelingen van een wetenschappelijke bibliotheek aangepakt. Voor het probleem wordt een optimalisatiemodel gedefinieerd met als doelstelling het maximaliseren van het gebruik van de digitale collectie over alle bibliotheekdivisies met een budgetrestrictie voor de volledige collectie. Door het gebruik van deze holistische benadering draagt dit onderzoek bij aan de wetenschap door middel van een geïntegreerde oplossing voor de bibliotheekmanagers. Het ontwikkelde systeem helpt economische beslissingen te nemen op basis van een "zo realistisch mogelijk" beeld en perspectief van de bibliotheeksituatie.

Resumen

Las bibliotecas, desde su creación hace más de 4000 años, han estado en un proceso de constante cambio. Aunque durante siglos estos cambios fueron muy lentos, en las últimas décadas, las bibliotecas universitarias han estado luchando continuamente para adaptar sus servicios a las siempre cambiantes necesidades de los estudiantes y personal académico. Además, la revolución del contenido digital, los avances tecnológicos, y los presupuestos cada vez más reducidos han obligado a las bibliotecas a asignar de manera eficiente sus limitados recursos entre su colección y servicios. Por desgracia, esta asignación de recursos es un proceso complejo debido a la diversidad de fuentes de datos y formatos que se requieren analizar antes de la toma de decisiones, así como la falta de métodos eficientes para su integración.

El objetivo principal de esta investigación es, a través de un análisis holístico, desarrollar un modelo integrado que ayude a las bibliotecas en la toma de decisiones óptimas sobre manejo de presupuestos y asignación de recursos entre sus servicios y colección. Con este fin, en este estudio se llevan a cabo una combinación de varias metodologías y enfoques estructurados. En primer lugar, se propone una estructura holística y el conjunto de herramientas necesarias para evaluar de manera integral las bibliotecas universitarias con el propósito de recopilar y organizar los datos desde el punto de vista económico. Se utiliza un marco teórico de cuatro cuadrantes en la que el sistema de la biblioteca y su colección son analizados desde la perspectiva de los usuarios y los actores internos. El primer cuadrante corresponde a la perspectiva interna del sistema de bibliotecas en el que se analiza el rendimiento de la biblioteca, los costos incurridos y los recursos consumidos por los servicios bibliotecarios. El segundo cuadrante evalúa la perspectiva externa del sistema de bibliotecas; es decir la percepción del usuario sobre la calidad de los servicios. El tercer cuadrante analiza la perspectiva externa de la colección de la biblioteca esto es evaluar el impacto de la colección de la biblioteca en sus usuarios. Finalmente, el cuarto cuadrante evalúa la perspectiva interna de la colección de la biblioteca; se analizan los patrones de uso utilizados para manipular la colección de la biblioteca.

Los datos para este marco de análisis proceden de múltiples fuentes y con diferentes formatos, por tanto, necesitan ser integrados y almacenados en un esquema adecuado para dar soporte a la toma de decisiones. Con este fin en segundo lugar, se ha diseñado e implementado una arquitectura de Data Warehouse que integra, procesa y almacena los datos recolectados a través del enfoque holístico. Finalmente, la información estratégica almacenada en el Data Warehouse es analizada y utilizada para diferentes propósitos, incluyendo los siguientes: 1) Visualización de datos y presentación de informes que permitan a los administradores de la biblioteca publicar sus indicadores de una manera sencilla y rápida. 2) Sofisticado análisis de la información mediante el uso de herramientas de minería de datos. En este estudio tres técnicas de minería de datos han sido examinadas: regresión, agrupamiento o clustering y clasificación. Estas técnicas de minería de datos se han aplicado al estudio de caso de la siguiente manera: predecir la futura inversión en el desarrollo de la biblioteca; búsqueda de grupos de usuarios que comparten intereses comunes y perfiles similares pero pertenecen a diferentes facultades; y predecir factores de la biblioteca que

pueden afectar e influenciar el rendimiento académico de los estudiantes. 3) Como entrada para la creación de modelos de optimización; en este estudio, se documentan las primeras experiencias de desarrollo de un modelo óptimo de asignación de recursos para distribuir recursos entre los diferentes procesos de un sistema de biblioteca. Al proponer un enfoque holístico, esta investigación contribuye al conocimiento, proporcionando una solución integrada que ayuda a los administradores de la biblioteca a tomar decisiones económicas basadas en una perspectiva "lo más realista posible" de la situación de la biblioteca.

List of Abbreviations

<i>3NF</i>	Third-Normal Form
<i>ABC</i>	Activity-Based Costing
<i>ACG</i>	Australian Competitive Grant
<i>AI</i>	Artificial Intelligence
<i>ALiLA</i>	Academic Library in Latin America
<i>ARIMA</i>	Autoregressive Integrated Moving Average
<i>ARL</i>	Association of Research Libraries
<i>BI</i>	Business Intelligence
<i>BSC</i>	Balanced Scorecard
<i>CBA</i>	Campusbibliotheek Arenberg – Arenberg Campus Library
<i>CDRJBV</i>	Regional Documentation Centre “Juan Bautista Vazquez”
<i>CM</i>	Computer Maintenance
<i>CONEA</i>	National Council for Accreditation of Higher Education – Ecuador
<i>COUNTER</i>	Counting Online Usage of NeTworked Electronic Resources
<i>CPA</i>	Customer Profitability Analysis
<i>CRM</i>	Customer Relationship Management
<i>DDC</i>	Dewey Decimal Classification
<i>DLA</i>	Deep Log Analysis
<i>DSS</i>	Decision Support System
<i>DW</i>	Data Warehouse
<i>ED</i>	Emergency Department
<i>ERP</i>	Enterprise Resource Planning
<i>ETL</i>	Extraction, Transformation and Load
<i>FOSS</i>	Free and Open Source Software
<i>FTE</i>	Full Time Equivalent
<i>iDSS</i>	Integrated Decision Support System
<i>IF</i>	Impact Factor
<i>ILL</i>	Inter-Library Loan
<i>ISU</i>	Iowa State University
<i>IT</i>	Information Technology
<i>JCR</i>	Journal Citation Reports
<i>KDD</i>	Knowledge Discovery in Databases
<i>KCGG</i>	Kenniscentrum voor de Gezondheidszorg Gent
<i>KM</i>	Knowledge Management
<i>GO</i>	General Overhead
<i>LIS</i>	Library and Information Science

<i>LISA</i>	Library and Information Science Abstracts
<i>LISTA</i>	Library, Information Sciences and Technology Abstracts
<i>LMS</i>	Library Management System
<i>LSSVM</i>	Least Squares Support Vector Machine
<i>OAI-PMH</i>	Open Archives Initiative Protocol for Metadata Harvesting
<i>OLAP</i>	Online Analytical Processing
<i>OLTP</i>	Online Transactions Processing
<i>OPAC</i>	Online Public Access Catalog
<i>ORBIL</i>	Optimal Resource allocation and Budgeting in Libraries
<i>PCA</i>	Principal Component Analysis
<i>RFID</i>	Radio Frequency Identification
<i>RM</i>	RFID Maintenance
<i>RMSE</i>	Root Mean Square Error
<i>RQ</i>	Research Question
<i>SCONUL</i>	Society of College, National and University Libraries
<i>SENESCYT</i>	National Secretary of Higher Education, Science, Technology and Innovation
<i>SERVQUAL</i>	SERvice QUALity
<i>SLE</i>	Student Library Employee
<i>SVM</i>	Support Vector Machine
<i>TDABC</i>	Time-Driven Activity-Based Costing
<i>TLA</i>	Transaction Log Analysis
<i>TQM</i>	Total Quality Management
<i>UC</i>	University of Cuenca
<i>UZ</i>	Universiteit Ziekenhuis
<i>VAT</i>	Value-Added Tax
<i>VLIR-IUC</i>	Vlaamse Interuniversitaire Raad- Institutional University Cooperation
<i>WBIB</i>	Wetenschappelijke BIBliotheek
<i>WEKA</i>	Waikato Environment for Knowledge Analysis
<i>WMAG</i>	Wetenschappelijke MAGazijn
<i>WoS</i>	Web of Science
<i>WWW</i>	World Wide Web

List of Contents

ACKNOWLEDGMENT I

ABSTRACT.....III

KORTE INHOUD..... V

RESUMENVII

LIST OF ABBREVIATIONS IX

LIST OF CONTENTS..... XI

LIST OF FIGURES..... XV

LIST OF TABLESXIX

PART I INTRODUCTION..... 1

CHAPTER 1: INTRODUCTION AND OVERVIEW3

1.1 Introduction 3

1.2 Theoretical Background..... 6

1.2.1 Decision Support Systems.....6

1.2.2 Holistic Approach for Data Collection7

1.2.3 Costing models8

1.2.4 Data Warehouse.....9

1.2.5 Data Mining.....9

1.2.6 Resource allocation and budgeting 10

1.3 Aims of the Project 11

1.3.1 Purpose Statement..... 11

1.3.2 Research Questions 11

1.4 Case Studies 12

1.5 Dissertation Overview 14

PART II DATA SOURCE 21

CHAPTER 2: HOLISTIC STRUCTURE APPROACH FOR DATA COLLECTION..... 23

2.1 Introduction 24

2.2 Theoretical background..... 25

2.3 Data collection through a holistic perspective 26

2.3.1 First quadrant: internal perspective of the library system..... 26

2.3.2 Second quadrant: external perspective of the library system 27

2.3.3 Third quadrant: external perspective of the library collection 29

2.3.4 Fourth quadrant: internal perspective of the library collection 30

2.3.5 The proposed holistic evaluation matrix 32

2.4 Data warehouse architecture for library holistic evaluation..... 33

2.4.1 Data source layer 33

2.4.2 Data extraction, cleansing and storage layer 34

2.4.3 Data presentation layer 34

2.5 Conclusions 34

CHAPTER 3: LITERATURE REVIEW OF TIME-DRIVEN ACTIVITY-BASED COSTING 41

3.1 Introduction 42

3.2	<i>Costing Systems</i>	42
3.2.1	Traditional costing systems	42
3.2.2	Activity-Based Costing System.....	43
3.3	<i>Time-Driven Activity-Based Costing System</i>	44
3.3.1	Brief history.....	44
3.3.2	The model.....	45
3.3.3	TDABC vs. ABC.....	48
3.3.4	Case studies	48
3.4	<i>Benefits and challenges</i>	55
3.4.1	Benefits.....	55
3.4.2	Challenges	56
3.5	<i>Opportunities</i>	58
3.6	<i>Conclusions</i>	59
CHAPTER 4:	TIME-DRIVEN ACTIVITY-BASED COSTING FOR LENDING AND RETURNING PROCESSES	65
4.1	<i>Introduction</i>	66
4.2	<i>Theoretical background: TDABC</i>	66
4.3	<i>TDABC in libraries</i>	68
4.4	<i>Case Study of Loan and Return Processes</i>	68
4.4.1	Step 1: Identifying the services or activities.....	68
4.4.1	Step 2: Estimating the total cost of resource groups.....	70
4.4.2	Step 3: Estimating the practical time capacity of each resource group	72
4.4.3	Step 4: Calculating the unit cost of each resource group	73
4.4.4	Step 5: Determining the time estimation for each activity.....	73
4.4.5	Step 6: Multiplying the unit cost of each resource group by the time estimate for the activity	74
4.5	<i>Implications</i>	75
4.6	<i>Conclusions</i>	79
CHAPTER 5:	TIME-DRIVEN ACTIVITY-BASED COSTING SYSTEMS TO MAXIMIZE PROCESS BENCHMARKING IN LIBRARIES	83
5.1	<i>Introduction</i>	84
5.2	<i>Research methodology</i>	85
5.3	<i>Results</i>	88
5.3.1	Acquisition	88
5.3.2	Cataloging.....	88
5.3.3	Circulation.....	90
5.3.4	Document Delivery	90
5.3.5	Overview.....	91
5.4	<i>Discussion</i>	93
5.4.1	Book acquisition process.....	94
5.4.2	Journal acquisition process	94
5.4.3	Requesting closed stack items process.....	95
5.5	<i>Conclusions</i>	96
PART III	DATA STORAGE	99
CHAPTER 6:	INTEGRATED DECISION-SUPPORT SYSTEM FOR LIBRARY HOLISTIC EVALUATION	101
6.1	<i>Introduction</i>	102
6.2	<i>Data collection through a holistic perspective</i>	102
6.3	<i>Research methodology</i>	103
6.4	<i>Designing an integrated decision support system: Case Study</i>	104
6.4.1	Decision support system architecture	104
6.4.1	Hefesto methodology	104
6.4.1	BLOCK 2: Data Integration.....	107
6.4.2	BLOCK 3: Data presentation.....	110
6.5	<i>Conclusions</i>	112
PART IV	DATA ANALYSIS AND PRESENTATION	115
CHAPTER 7:	LITERATURE REVIEW OF DATA MINING APPLICATIONS IN LIBRARIES.....	117
7.1	<i>Introduction</i>	118

7.2	<i>Research methodology</i>	119
7.3	<i>Classification method</i>	122
7.3.1	Classification framework: Holistic approach for library evaluation.....	122
7.3.2	Classification framework: Data mining models.....	122
7.3.3	Research classification process	123
7.4	<i>Classification of the articles</i>	123
7.4.1	Distribution of articles by the library holistic quadrants and data mining models 124	
7.4.2	Distribution of articles by year of publication	128
7.4.3	Distribution of articles by country of implementation.....	132
7.4.4	Distribution of articles by journal in which the articles were published	132
7.4.5	Distribution of articles by library type analyzed	133
7.5	<i>Limitations</i>	133
7.6	<i>Conclusion and research implications</i>	134
CHAPTER 8:	AN EXPERIMENTAL USE OF BIBLIOMINING	139
8.1	<i>Introduction</i>	140
8.2	<i>Theoretical Background</i>	140
8.2.1	Holistic data collection	140
8.2.2	Data Warehouse.....	141
8.2.3	Bibliomining.....	142
8.3	<i>Related Work</i>	143
8.4	<i>Data Warehouse methodology selection</i>	144
8.5	<i>iDSS for library holistic evaluation</i>	144
8.5.1	Architecture of the iDSS.....	145
8.5.2	Design of the iDSS.....	145
8.6	<i>Experimental use of bibliomining for library decision-making</i>	146
8.6.1	Predicting the future investment in library development	146
8.6.2	Users clustering for selective dissemination of information.....	151
8.6.3	Predicting factors that affect student academic performance.....	157
8.7	<i>Conclusions</i>	162
CHAPTER 9:	TOWARDS AN OPTIMIZATION MODEL FOR RESOURCE ALLOCATION IN LIBRARIES	167
9.1	<i>Introduction</i>	168
9.2	<i>Theoretical Background</i>	169
9.2.1	Holistic Approach for Data Input	169
9.2.2	Resource allocation	170
9.2.3	Budgeting.....	171
9.3	<i>Budgeting in university libraries</i>	171
9.4	<i>Towards a resource allocation model</i>	174
9.4.1	Case Study	174
9.4.2	Usage analysis.....	174
9.4.3	Formulating the model.....	177
9.5	<i>Conclusions</i>	179
PART V	CONCLUSIONS.....	183
CHAPTER 10:	CONCLUSIONS.....	185
10.1	<i>Introduction</i>	185
10.2	<i>Summary of significant findings with respect to the formulated research propositions</i> 185	
10.2.1	Data Source.....	186
10.2.2	Costing analysis.....	187
10.2.3	Data Storage	189
10.2.4	Data Analysis and Presentation.....	190
10.3	<i>Optimal Resource allocation and Budgeting in Libraries. The ORBIL approach</i>	192
10.4	<i>Overall conclusion</i>	206
APPENDICES	207
APPENDIX A:	TOWARDS A HOLISTIC ANALYSIS TOOL TO SUPPORT DECISION-MAKING IN LIBRARIES.....	209
A.1	<i>Introduction</i>	210

A.2 Theoretical background.....	210
A.3 Holistic analysis tool to support decision-making in libraries: Case study.....	211
A.3.1 The case study.....	211
A.3.2 First quadrant: internal perspective of the library system.....	211
A.3.3 Second quadrant: external perspective of the library system	213
A.3.4 Third quadrant: external perspective of use	214
A.3.5 Fourth quadrant: internal perspective of use	215
A.4 Conclusion.....	215
APPENDIX B: TIME-DRIVEN ACTIVITY-BASED COSTING SYSTEMS FOR CATALOGING PROCESSES.....	217
B.1 Introduction.....	218
B.2 Theoretical background: Costing systems.....	218
B.2.1 Traditional Costing Systems.....	218
B.2.2 Activity-Based Costing Systems	219
B.2.3 Time-Driven Activity-Based Costing Systems	219
B.3 TDABC in Cataloging Processes	220
B.3.1 Original Cataloging.....	223
B.3.2 Copy Cataloging.....	224
B.4 Original and Copy Cataloging.....	225
B.5 Benefits of TDABC in Cataloging Processes.....	232
B.6 Conclusions.....	233
APPENDIX C: TD-ABC-D: TIME-DRIVEN ACTIVITY-BASED COSTING SOFTWARE FOR LIBRARIES	237
C.1 Introduction.....	237
C.2 Poster	238
APPENDIX D: USING TIME-DRIVEN ACTIVITY-BASED COSTING TO IDENTIFY BEST PRACTICES IN	
ACADEMIC LIBRARIES: COSTING TABLES.....	241
D.1 Introduction.....	241
APPENDIX E: REFERENCES OF THE EMPIRICAL STUDIES IMPLEMENTING DATA MINING TECHNIQUES IN	
LIBRARIES 251	
E.1 Introduction.....	251
CURRICULUM VITAE.....	255
LIST OF PUBLICATIONS.....	257

List of Figures

Figure 1.1: Structure of Chapter 1..... 6

Figure 1.2: Conceptual Matrix for Holistic Measurement (Nicholson, 2004)..... 7

Figure 1.3: Structure of the dissertation 14

Figure 2.1: Conceptual matrix for holistic measurement (Nicholson, 2004) 24

Figure 2.2: Methodologies proposed to economical evaluate a library through a holistic perspective 32

Figure 2.3: Data warehouse architecture for library holistic evaluation 33

Figure 3.1: Activity-Based Costing Structure (Kaplan & Cooper, 1998) 43

Figure 3.2: TDABC model (based on (Everaert, Bruggeman, Sarens, et al, 2008)) 46

Figure 4.1: WBIB Lending Process 69

Figure 4.2: WMAG Lending Process..... 71

Figure 4.3: Returning Process..... 72

Figure 5.1: Percentage of time monthly consumed per process 92

Figure 5.2: Percentage of cost monthly consumed per process 92

Figure 5.3: Process comparison based on time and cost indicators: (a) Time and (b) Cost..... 93

Figure 5.4: Book acquisition process comparison based on time and cost indicators 94

Figure 5.5: Journal acquisition process comparison based on time and cost indicators 95

Figure 5.6: Requesting closed stack items process comparison based on time and cost indicators . 96

Figure 6.1: Methodologies proposed for the economic evaluation of libraries through a holistic approach (Siguenza-Guzman et al., 2015)102

Figure 6.2: iDSS architecture of the UC library.....105

Figure 6.3: Indicators and perspectives of the lending process example.....105

Figure 6.4: Conceptual data model of the UC library data warehouse.....108

Figure 6.5: Logical model of the lending process example.....108

Figure 6.6: Multidimensional model of the lending process.....109

Figure 6.7: Evaluation of string similarity in the personal author field110

Figure 6.8: Number of loans of a particular item, of a given author and title, of a specific item category of a particular librarian.....111

Figure 6.9: Number of loans of a particular librarian of a specific library branch.....111

Figure 6.10: Operating costs of lending services of a particular campus112

Figure 7.1: Data mining process, based on Han et al. (2011).	118
Figure 7.2: Classification framework for data mining techniques based on the Ngai et al. (2009) approach.	121
Figure 7.3: Selection process framework.	124
Figure 7.4: Classification of data mining applications based on the holistic evaluation matrix	129
Figure 7.5: Evolution of number of articles per year	130
Figure 7.6: Chronological evolution of articles by library holistic evaluation. (a) Service analysis; (b) Quality analysis; (c) Collection analysis; (d) Usage analysis.	131
Figure 7.7: Distribution of articles by country of implementation.	132
Figure 7.8: Distribution of articles in the top 6 journals	132
Figure 7.9: Distribution of articles by type of library analyzed	133
Figure 8.1: Methodologies proposed for the economic evaluation of libraries through a holistic approach (Siguenza-Guzman, Van Den Abbeele, et al., 2015).	141
Figure 8.2: Architecture of a DW (Based on Inmon (2005)).	142
Figure 8.3: The Hefesto methodology.	145
Figure 8.4: iDSS architecture of ALiLA	145
Figure 8.5: Acquisition values 2000-2014. (a) Original and cleaned time series after applying an outlier replacement process. (b) Cleaned and smoothed time series after applying a noise removal filter.	147
Figure 8.6: Out-of-sample evaluation.	148
Figure 8.7: RMSE results of the three predictive models: (a) SVM, (b) LSSVM and (c) ARIMA	150
Figure 8.8: Comparison of the three techniques using step-ahead forecasting for the test period 2011 – 2014	150
Figure 8.9: Expenses behavior: (a) per year; (b) per month. Horizontal dashed line represents the mean values	151
Figure 8.10: (a) Total number of users per faculty; (b) Number of user transactions per faculty	152
Figure 8.11: Silhouette score of the clustering algorithms at different configurations	153
Figure 8.12: Cluster membership grouped according to faculty of (a) spectral clustering (n=2); (b) mini-batch k-means (n=16); and, (c) ward (n=20)	154
Figure 8.13: Statistics of the cluster 14 produced by ward (n=20): a) number of users per faculty; b) income; c) loan duration; d) academic performance.	155
Figure 8.14: User behavior in terms of topics of interest per faculty for cluster 14, ward (n=20); (a) user interests in terms of categories; (b) user interests in terms of subcategories	156
Figure 8.15: Statistics of the full dataset: (a) income; (b) loan duration; (c) academic performance	158
Figure 8.16: High dimensional visualization of the complete dataset. (a) Parallel coordinates. (b) Andrew's curves	159
Figure 8.17: Statistics of the career of Medicine and Surgery: a) income; b) loan duration; c) academic performance	160
Figure 8.18: High dimensional visualization of the Medicine dataset. (a) parallel coordinates. (b) Andrews curves.	161

Figure 9.1: Methodologies proposed for the economic evaluation of libraries through a holistic approach (Siguenza-Guzman et al., 2015)	170
Figure 9.2: Structure classification used for determining the weights	176
Figure 9.3: Exemplification of the resource allocation problem	178
Figure 10.1: Methodologies proposed for the economic evaluation of libraries through a holistic approach.....	186
Figure 10.2: iDSS architecture of the CDRJBV.....	190
Figure 10.3: ORBIL Framework	193

Appendices

Figure A.1: Conceptual matrix for holistic measurement (Nicholson, 2004).....	210
Figure B.1: Cataloging Processes: a) Original Cataloging; b) Copy Cataloging	221
Figure B.2.: Resulting activity flow of the Cataloging Process	224
Figure B.3: Comparison of original and copy cataloging in terms of time and cost.....	228
Figure B.4: Pareto chart of disaggregating costs per activity for original and copy cataloging	229
Figure B.5: Improvements on the cataloging process	232

List of Tables

Table 3.1: Case studies sorted by area..... 49

Table 4.1: Time-Driven Activity-Based Costing steps (Everaert et al., 2008) 67

Table 4.2: Unit cost per resource group..... 73

Table 4.3: WBIB Lending Process Cost Table 76

Table 4.4: WMAG Lending Process Cost Table 77

Table 4.5: Returning Process Cost Table 78

Table 4.6: Returning Process through the Chico Robot (Three Items Returned)..... 79

Table 4.7: Returning Process through the Hand-in-Robot (Three Items Returned)..... 79

Table 4.8: Returning Process through the librarian staff..... 79

Table 5.1: Processes to be analyzed using TDABC..... 86

Table 5.2: Costs involved in the analysis 87

Table 5.3: Books acquisition process cost table (Library 1)..... 89

Table 5.4: Books acquisition process cost table (Library 2) 89

Table 5.5: TDABC Cost Benchmarks between library 1 and library 2 92

Table 6.1: Summary of data sources of the UC library107

Table 7.1: Search criteria and number of results per database121

Table 7.2: Distribution of articles according to the proposed classification model125

Table 7.3: Distribution of articles by the library holistic quadrant and data mining models127

Table 7.4: Distribution of articles by data mining techniques and library holistic quadrants129

Table 7.5: Distribution of articles by year of publication and country of implementation.....130

Table 8.1: Average RMSE results per forecaster.....150

Table 8.2: Variables used for clustering.....152

Table 8.3: Variables used for classification.....157

Table 8.4: Correlation of variables including their p-value.....158

Table 8.5: Classification results for the training set.....160

Table 8.6: Classification results for the testing set.....160

Table 8.7: Confusion Matrix for the binary classification model on the testing set.....160

Table 8.8: Correlation of variables including their p-value.....161

Table 8.9: Classification results for the testing set.....161

Table 8.10: Confusion Matrix for the Binary Classification Model on the training set162

Table 9.1: Definition, advantages and disadvantages of the budget system approach (Linn, 2007; Blake Gonzalez, 2011).....172

Table 10.1: Description of the first block, including activities, deliverables, best practices, indicators and future work.....	194
Table 10.2: Example of the first block of the ORBIL approach applied to books.....	199
Table 10.3: Example of the first block of the ORBIL approach applied to Circulation and ILL services	200
Table 10.4: Description of the second block, including activities, deliverables, best practices, indicators and future work.....	201
Table 10.5: Description of the third block, including activities, deliverables, best practices, indicators and future work.....	204

Appendices:

Table A.1: Processes analyzed using TDABC.....	212
Table A.2: Action points after LibQUAL+® 2012	214
Table B.1: Time-Driven Activity-Based Costing steps (Everaert et al. 2008).....	221
Table B.2: Total cost of each resource group	222
Table B.3: Costs involved in the analysis	223
Table B.4: Total cost of the cataloging process	226
Table B.5: Cataloging costs adjusted using inflation and exchange-rates	227
Table B.6: Example of a what-if analysis applied to the original model of the cataloging process...230	
Table B.7: Example of a what-if analysis to improve applied to the original model of the cataloging process	231
Table D.1: Journals acquisition process cost table (Library 1)	242
Table D.2: Journals acquisition process cost table (Library 2)	242
Table D.3: Original cataloging process cost table (Library 1)	243
Table D.4: Original cataloging process cost table (Library 2)	243
Table D.5: Copy cataloging process cost table (Library 1)	244
Table D.6: Copy cataloging process cost table (Library 2)	244
Table D.7: Lending process cost table (Library 1)	244
Table D.8: Lending process cost table (Library 2)	245
Table D.9: Returning process cost table (Library 1)	245
Table D.10: Returning process cost table (Library 2)	245
Table D.11: Requesting closed stack items process (Library 1).....	246
Table D.12: Requesting closed stack items process (Library 2).....	246
Table D.13: ILL outgoing request process (Library 1)	247
Table D.14: ILL outgoing request process (Library 2)	247
Table D.15: ILL incoming request process for digital items (Library 1)	248
Table D.16: ILL incoming request process for digital items (Library 2)	248
Table D.17: ILL incoming request process for printed items (Library 1).....	248
Table D.18: ILL incoming request process for printed items (Library 2).....	249

PART I

Introduction

Chapter 1: Introduction and Overview

“A child, a teacher, a book and a pencil could change the world”

Malala Yousafzai

Nobel Peace Prize 2014

1.1 Introduction

Libraries have been present in our society for about 4000 years. The first libraries originated as archives in ancient Egypt and Mesopotamia, and consisted merely of archives of clay tablets, papyrus and other writing surfaces (Casson, 2002). These tablets contained detailed information and knowledge of the economic life of communities and of their governments. Usually, libraries at that time were not open to the public but to the exclusive use of priests and rulers. Librarians, in turn, are as old as libraries themselves; their presence has been illustrated by the evidence of catalogs found in some destroyed ancient libraries (Mukherjee, 1966).

In the Hellenistic world, the Library of Alexandria in Egypt is considered to be the first attempt to bring all human knowledge together in only one place. Although this library was not the first effort at forming a library, it was considered as one of the largest and most significant great libraries of its period (Trumble, 2003). In the early Middle Ages, the most important libraries in Europe were attached to cathedrals and monasteries, such as the Abbey of Montecassino in Italy (Setton, 1960), where copies of manuscripts were hand-created and disseminated by monks (Byfield, Project, & Stanway, 2004). In the late Middle Ages, on the contrary, libraries were more abundant, mainly because several libraries were created to serve teaching and research like in the universities of Paris and Oxford (Mugridge, 2012).

Libraries played an important role in the Renaissance and Reformation, and then in the Enlightenment period. In fact, with the invention and spread of the printing press, in the middle of the fifteenth century, books became more accessible due to the increase in the number of issues and circulation books (Uhlendorf, 1932). The 17th and 18th centuries were considered the golden age of libraries (Enlightenment era), when most of the book collections were begun and libraries surged in popularity (Stockwell, 2000). During this period, as public universities developed, very important libraries grew as well as national libraries. At the start of the 18th century, libraries were becoming increasingly public and were more frequently lending libraries. The nineteenth century is particularly important because it is the century in which public libraries were widespread; first legislations were enacted and the first professional library associations were begun.

Libraries were historically known for their gatekeeping functions, preservation expertise, and information provider services, where libraries kept paper copies of books and journals, and

provided the appropriated tools and guidance on how to access their own resources and resources from other libraries (Kennan & Wilson, 2006). After centuries of stationary growth, libraries in the 21st century faced enormous challenges. Within a quarter of a century, libraries moved from card catalogs to online public access catalogs (OPACS), from printed indexes to CD-ROMs, and from CD-ROMs to Web-based databases that can be searched remotely (Stephens, 1998; Lynch, 2000). Libraries, that for centuries were fulfilling users' needs in a traditional way, experienced a revolutionary and exponential change in the manner in which they provided their offered services. In the last decades, academic libraries, in particular, have passed through three automation ages (Lynch, 2000). The first automation age for most academic libraries, reaching back to the 1960s and early 1970s, began with the computerization of library processes. Mini-computers were introduced to barcode books and to automate manual processes such as circulation and cataloging. Lynch (2000) highlights the development of shared copy-cataloging systems as the greatest achievement of this period. By the 1980s and early 1990s, the second automation age started with the rise of public access. Academic libraries were confronted with environmental changes driven by information technology, which quickly moved the focus of attention away from automation toward a series of much more fundamental questions about library roles and missions such as new roles of consortia and business models. In fact, resource sharing had major investments in this period; consortia developed union catalogs that merged serials in a journal level, and computer-assisted interlibrary loan systems were built on the shared national union catalog databases. In addition, Lynch points out that the greatest financial success was achieved by creating collective purchasing consortia to negotiate prices for all members of the consortium. An additional result of this second round of changes was the OPAC as a replacement for the traditional card catalogs. Users could search library holdings at any time and even remotely, rather than having to go to the library. Finally, the last and more challenging automation age comes with the transition from print to digital content.

Libraries, since the last years of the twentieth century, have been in a process of constant change, and academic libraries in particular have strived to adapt their services to the ever-changing user needs of their students, researchers, and academics (Alvite & Barrionuevo, 2010; Delaney & Bates, 2014; Ellis, Rosenblum, Stratton, & Ames-Stratton, 2014). Inevitably, users have changed their information-seeking behavior due to rapid technological advances and astonishing e-content revolution; the growing presence of e-books and the proliferation of tablets and mobile devices have transformed the manner how information is disseminated and consumed (Allen Press, Inc., 2012; Bertot, 2011; Brook & Salter, 2012). This situation has obliged libraries to automatize or digitalize their current services, redevelop facilities and physical infrastructure, as well as promote new initiatives such as digital services and organizational structures.

Libraries have been adapting their traditional services, driven by technological developments, and increases in the use of technology. For instance, with the widespread use of the Internet and search engines, such as Google, users have little or no problem finding information on Web sources; consequently, the use of OPAC is steadily declining (ACRL Research Planning and Review Committee, 2012; Danskin, 2007; Ross & Sennyey, 2008). In the specific case of academic libraries, the emergence of new digital technologies has altered traditional libraries beyond recognition (Raju, 2014). As new technologies and information delivery systems emerge, the way in which students, researches, and academics search for information is also changing (Nicholas, Huntington, Jamali, Rowlands, & Fieldhouse, 2009; Niu et al., 2010). Examples of technological developments transforming the nature of academic and research environment include the following: growing popularity of massive online open courses; explosive use of mobile devices; development of globally-networked research communities; as well as growth of new pedagogical methods, including flipped classrooms, online and distance learning, experiential and project-based learning, and student-centered research (Delaney & Bates, 2014).

Libraries, motivated by the growing presence of e-content and the new modes used by people to access information in electronic resources, have been adapting their traditional collection and have been creating new digital content as future promises an increased amount of digital information. Furthermore, the recent emphasis on open access, open data, data-plan management, and "big data" research are creating the impetus for academic institutions to develop and deploy new initiatives,

service units, and resources to meet scholarly needs at various stages of the research process (ACRL Research Planning and Review Committee, 2014). However, to facilitate access to these e-resources, academic libraries are dealing with several challenges such as the lack of uniformity in license terms, lease conditions, access restrictions, and librarians' expectations (Walters, 2013). Furthermore, e-services like remote access to digital information has meant that many students, researchers and professors misunderstand what libraries do for them and do not necessarily associate the library with providing information resources. Therefore, in this current evolving information environment, libraries recognizing that their intrinsic information provider role is becoming less and less visible, have responded to these challenges and technological developments by rethinking and repurposing what libraries are and what libraries do for their users (Delaney & Bates, 2014). In fact, although libraries have been present in our society for centuries, authors such as Borgman (2003) think that they are becoming invisible because everything that users need can be found online.

Libraries in general have been threatened for many years by stagnant or shrinking budget constraints driven by global financial crisis (Sudarsan, 2006a; McKendrick, 2011; Guarria & Wang, 2011). This situation stems from library services usually perceived as "free of charge" but in reality, not free of costs, which strongly depends on institutional funding (Stouthuysen, Swiggers, Reheul, & Roodhooft, 2010). Moreover, although the migration of physical to digital environments has facilitated managing information and allowed access to a number of digital journals and e-books, it has also contributed to the escalating collection costs, as well as an increase in the complexity of budgeting models and resource allocation processes (Chan, 2008; Guarria, 2009; Poll, 2001). For instance, one of the problems with a subscription-based digital library collection and patron-driven acquisition is the variability of their yearly prices, which has rapidly risen in the last years (Allen Press, Inc., 2012). In fact, the most alarming trend in the academic library environment is the increase of information resource expenditures (Blake Gonzalez, 2011). Chan (2008) affirms that digital resource expenditures had a yearly average growth of 25%, while library budgets only had an average growth of 2.3%. These economic constraints result in a tremendous financial pressure for library directors, whom are required to shift budgeting and spending priorities (Blake Gonzalez, 2011). As a consequence, several decisions have been made such as cutting collection budgets, eliminating budgets for travels or conferences, freezing salaries, and finding new ways to fund programs (Sudarsan 2006, McKendrick 2011).

All these funding constraints, as well as technological developments that can be seen as opportunities or potential challenges, are proof that libraries have to be more innovative in providing, justifying, and evaluating the efficiency and effectiveness of their services. Libraries more than ever must evolve and continue to demonstrate their relevance to the academic management, who faces difficulties understanding the new roles, cost, and value of good libraries (ACRL Research Planning and Review Committee, 2012, 2013). To do so, libraries have increased their focus on assessment of outcomes over inputs and placed emphasis in demonstrating that these outcomes are having an impact on academic libraries and parent institutions. Additionally, libraries are also increasing their understanding of their users, collection and services, and related costs in order to justify resource requirements. Because of limited funding, library administrators are assessing the best ways to allocate their resources, how to redefine themselves, and reengineer their budget strategies.

Resource allocation with limited money, staff, and infrastructure between library services and collection is a complex process due to the high number of constraints and data sources that require consultation prior to decision-making, as well as due to the lack of efficient integration methods. Although many resource allocation approaches have been developed, most of them have mainly focused on the distribution of money for either physical or digital collections. There is also a lack of awareness in embracing different perspectives from heterogeneous stakeholders such as researchers, developers, administrators, librarians, and general users (Zhang, 2010). Therefore, scientific approaches on how to allocate limited resources among shifting collections and dynamic services become crucial for libraries. This introductory chapter continues by presenting an overview of the main topics covered during the research: aims of the project, theoretical background and case studies analyzed. Figure 1.1 illustrates the structure of this chapter.

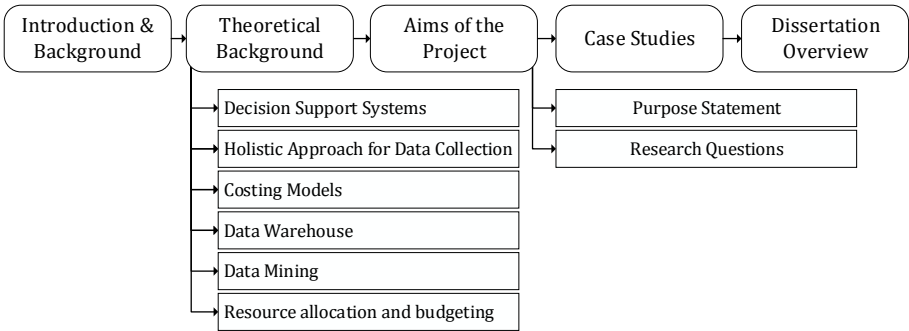


Figure 1.1: Structure of Chapter 1

1.2 Theoretical Background

In this section, the theoretical framework of the research study is outlined with a summary of the literature review on the main aspects covered: decision support systems, holistic approach for data collection, costing models, data warehouse, data mining, resource allocation, and budgeting.

1.2.1 Decision Support Systems

Libraries have been facing in the last decades significant challenges. The current dynamic library landscape, caused by limited budgets, rapid technological developments, and astonishing e-content growth, highlights the importance of understanding financial management in academic libraries (Blake Gonzalez, 2011). Strategic decision-making becomes essential in the allocation process of limited resources. Nevertheless, this type of decision-making process in academic libraries is highly complicated due to the large number of data sources, processes, and high volumes of data to be analyzed. Typical data sources include integrated library systems, library portals and OPACs, systems of consortiums, quality surveys, and university management systems. These heterogeneous data sources are only partially used for decision-making processes due to the wide variety of formats, standards and technologies, as well as the lack of efficient integration methods.

Traditional library management systems have been used for decades in meeting the automation needs of print-based libraries. Yet, traditional systems provide very limited functions and cannot promote innovation and knowledge creation for strategic decision-making. In this sense, knowledge management (KM) has become a powerful tool for libraries to expand their role and responsibilities to areas where they had little impact in the past, such as in financial decisions and strategic decision-making (Hobohm, 2004; Townley, 2001).

Knowledge-based Decision Support Systems (DSS) provide important information for library decision-making and performance improvement (Lai, Wang, Huang, & Kao, 2011). A typical knowledge discovery process is an interactive sequence of steps that normally starts by cleaning the data in order to remove noise, duplicate and inconsistent data (Han, Kamber, & Pei, 2011). Then, these cleaned data are integrated from multiple data sources and formats. Relevant data are selected and collected from the databases as raw data to be mined; these raw data are then transformed into appropriated formats that can be understood by other tools such as data mining or optimization models, and applied data filtration and aggregation techniques to integrate data from multiple tables into a single table. Interesting knowledge is extracted from the transformed data (or Knowledge Discovery in Databases – KDD), and this information is analyzed in order to identify truly interesting patterns. Eventually, strategic knowledge is visualized to managers to support decision-making.

Several DSS have been documented in literature to support decision making in libraries; however, most of them mainly focus on specific areas such as distribution of money for physical or digital collections, performance assessment of the library collection, and analysis of user behavior. Little is known about integrating all these different aspects and incorporating others such as human resources, technological infrastructure, processes performance, or usage indicators. Even less is known about embracing different perspectives from heterogeneous stakeholders such as researchers, academics, managers, librarians, and general users (Zhang, 2010). Hence, a crucial stage to the success of DSS implementation is the data collection process.

1.2.2 Holistic Approach for Data Collection

Ernst and Segall (1995) state that institutions with limited budgets and with difficult circumstances are called to develop strategic and well-coordinated budgeting plans by means of “holistic approaches”. The objective of a holistic approach is to help organizations to define a set of measures that reflect their objectives and assess their performance appropriately (Matthews, 2011). This holistic approach requires interconnecting all necessary components to evaluate the impact of limited resources in the whole institution, and then prioritize and optimize resource allocation in library services and collection. Many resource allocation approaches have been proposed; however, most of them mainly focus on economic allocation for either physical or digital collections separately. To the extent of our knowledge, the most complete approach to evaluate libraries from a holistic perspective is given by Nicholson (2004). The author proposes a theoretical analysis framework to support libraries in gaining a more thorough and comprehensive understanding of their users and services for both digital and physical services. This theoretical analysis framework is based on a two-dimensional evaluation matrix in which columns represent the topic, library system and use, and rows represent the perspective of the library system and users. An overview of the conceptual matrix for holistic measurement is shown in Figure 1.2.

	<i>Library System</i>	<i>Use</i>
<i>Internal Perspective (Library System)</i>	1. What does the library system consist of?	4. How is the library system manipulated?
<i>External Perspective (User)</i>	2. How effective is the library system?	3. How useful is the library system?

Figure 1.2: Conceptual Matrix for Holistic Measurement (Nicholson, 2004)

The main characteristics of each quadrant proposed by Nicholson (2004) are given hereinafter. The first quadrant corresponds to the internal perspective of the library system. This is a traditional type of analysis that can include bibliographic collection aspects, organizational flows, computer interfaces, processes, staff, and resources. The second quadrant evaluates the external perspective of the library system. User's perception about service quality is judged in this quadrant. Aboutness, effectiveness, and usability of the library services are the main aspects studied. The third quadrant analyses the external perspective of the library collection. This quadrant allows quantification of the impact of the library collection on its users, thus providing library managers with better basis for decision making when acquiring new bibliographic materials. By evaluating the current bibliographic collection, libraries may discover possible gaps and plan future collection development (Agee, 2005). Eventually, the fourth quadrant evaluates the internal perspective of the library collection. The interaction that a user had with the library system is analyzed. For instance, in digital library services, unlike circulation patterns in traditional services, it is possible to track everything users do within the library system, allowing libraries not only to know what users retrieve but also what they looked for and could not receive. This theoretical framework, thereby, requires retrieving and integrating information from various separate sources in order to be used in an adequate holistic library measurement.

1.2.3 Costing models

Academic libraries are accustomed to producing and gathering a vast amount of statistics about their collection and services; however, service and process costs are rarely calculated as a performance measure. Libraries have been the least cost-effective providers due to their unaccustomedness to perform formal costing analysis to their services and processes (E. Stewart Saunders, 2003); traditional costing systems being their most widely used technique. In a traditional costing system, direct costs such as direct labor and materials, are directly attributed to the services. On the contrary, indirect costs such as marketing, depreciation, training, and electricity are typically allocated to each service using a single or a few volume-based cost drivers (e.g., direct labor, service hours, or units of output). Traditional costing systems are adequate when indirect expenses are low and service variety is limited (Ellis-Newman & Robinson, 1998). However, in environments with a broad range of services, such as libraries, indirect costs have increasingly become more important than direct costs.

Seeking to remedy these limitations, libraries started employing more advanced cost calculation techniques, such as activity-based costing (ABC). ABC is an alternative costing system promoted by Cooper and Kaplan (1988). Compared to traditional costing methods, ABC performs a more accurate and efficient treatment of indirect costs (Ellis-Newman & Robinson, 1998). ABC first accumulates overhead costs for each activity, and then assigns the costs of the activities to the services causing that activity. An activity for libraries is defined as an event or task undertaken for a specific purpose such as cataloging, loan processing, shelving, and acquisition orders (Ellis-Newman, 2003). An extensive stream of literature describes ABC as a system that provides interesting advantages to decision-making in libraries (Ching, Leung, Fidow, & Huang, 2008; Ellis-Newman, 2003; Ellis-Newman & Robinson, 1998; Gerdson, 2002; Goddard & Ooi, 1998; Heaney, 2004; Novak, Paulos, & Clair, 2011; Skilbeck & Connell, 2001). However, ABC has great limitations: for instance, a high degree of subjectivity involved in estimating employees' proportion of time spent on each activity; the excessive time, resources and money for data collection; and the difficulties to model multi-driver activities (Dalci, Tanis, & Kosan, 2010; Demeere, Stouthuysen, & Roodhooft, 2009; Everaert, Bruggeman, & De Creus, 2008; Everaert, Bruggeman, Sarens, Anderson, & Levant, 2008; Kaplan & Anderson, 2004, 2007a; Tse & Gong, 2009; Wegmann & Nozile, 2009).

Time-Driven Activity-Based Costing (TDABC) is a costing system technique developed by Kaplan and Anderson to overcome the limitations of former approaches (Kaplan & Anderson, 2003). Hence, the majority of the TDABC advantages are based on the weaknesses of former approaches (Dejnega, 2011). TDABC uses only two parameters to assign resource costs directly to the cost objects: 1) the unit cost of supplying resource capacity; and 2) an estimated time required to perform an activity (Kaplan & Anderson, 2007b). For each activity, costing equations are calculated based on the time required to perform an activity. This time can be readily observed, validated, and then computed by time equations which are the sum of individual activity times (Kaplan & Anderson, 2007b). By using these equations, all possible combinations of activities can be represented for example, when different types of services do not necessarily require the same amount of time to be performed. Up to now, few case studies have been implemented in libraries regarding very specific processes such as the inter-library loan (Pernot, Roodhooft, & Van den Abbeele, 2007) and acquisition processes (Stouthuysen et al., 2010). In these case studies, TDABC is described as a tool that offers a relatively quick and cost-efficient way to design useful costing models, as well as to provide accurate information of the library activities which may help managers to get a better understanding of how the library uses its time, costs, and resources. These initial investigations show promising possibilities of using TDABC to provide accurate information on library activities. However, these studies have been applied to very specific settings, studying only particular processes with cases in small and medium-size academic libraries. More research is still needed to identify whether the technique of TDABC is useful and feasible to implement for a more extensive set of library activities.

1.2.4 Data Warehouse

William H. Inmon, acknowledged as the father of the “Data Warehouse” (DW), defines a DW as a “subject-oriented, integrated, time-variant and non-volatile collection of data in support of management’s decision making process” (Inmon, 2005, p. 29). According to Ralph Kimball (2006), an other preeminent figure in data warehousing, a DW is a repository of integrated information from distributed, autonomous, and possibly heterogeneous sources, specifically, structured for analysis and consultation. A DW is a read-only analytical database that integrates all information from various operational data sources whose purpose is to generate reports and analyze data in order to support the strategic decision-making process in an enterprise.

The aim of a DW is to consolidate the information locked up in the different operational databases with information from other, often external data sources, and make it available for data analysis from a managerial perspective (Singh, 1998). Data stored in the DW are snapshots resulting from data transformations, quality control checks, and integration of operational data. The major benefit of a DW is the possibility to have interactive and immediate access to strategic information of an enterprise. Users with a managerial role in the organization make their own inquiries and crossing data, using specialized tools with graphical interfaces such as data mining and on-line analytical processing tools. Operations in a DW are not predominant transactions, as in operational databases but complex queries of joining, filtering, grouping, and aggregating large amounts of data (Wrembel & Koncilia, 2007).

Due to the special characteristics of a DW, the design strategies used for building and managing operational databases are generally not applicable to the design of data warehouses (Inmon, 2005; Kimball, 2006). The design of a DW is an inherently complex, costly, and time-consuming task. Building a DW involves extracting data from different data sources, in which many problems of inconsistency need to be dealt with (Ying Wah, Hooi Peng, & Sue Hok, 2007). Data extraction, cleansing and storage through ETL (Extract, Transform, Load) processes is also a complex and time-consuming issue because this process needs to combine all the different data sources and convert them into a uniform format, excluding possible inconsistencies, redundancies, and incompatibilities (Nicholson, 2003). Therefore, to ensure the end result meets its needs, successful implementation of a DW requires, among other things, a significant investment of time and energy of those who will be its users, identifying the business objectives, designing the DW architecture and implementing the DW system (Curtis & Joshi, 2011; Radulescu, 2003). However, as Nicholson and Stanton (2009) remark, only by combining and linking different data sources, library managers can uncover the hidden strategic information that can help to properly understand processes and services in library decision-making.

1.2.5 Data Mining

Data mining, also known as knowledge discovery in databases (KDD), can be defined as the process of analyzing large information repositories and of discovering implicit, but potentially, useful information (Han et al., 2011). Data mining has the capability to uncover hidden relationships and to reveal unknown patterns and trends by digging into large amounts of data (Sumathi & Sivanandam, 2006). The functions (or models) of data mining can be categorized according to the task performed: association, classification, prediction, and clustering (Hui & Jha, 2000; Kao, Chang, & Lin, 2003; Nicholson, 2006b).

Data mining analysis is based normally on three methods: classical statistics, artificial intelligence, and machine learning (Girija & Srivatsa, 2006). Classical statistics is mainly used for studying data, data relationships and for dealing with numeric data in large databases (Girija & Srivatsa, 2006; Hand, 1998). Examples of classical statistics include regression analysis, cluster analysis, and discriminate analysis. Artificial intelligence (AI) is used for applying “human-thought-like” processing to statistical problems (Girija & Srivatsa, 2006). AI uses several techniques such as genetic algorithms, fuzzy logic, and neural computing. Finally, the last technique is machine learning that can be described as a combination of advanced statistical methods and AI heuristics,

used for data analysis and knowledge discovery (Kononenko & Kukar, 2007). Machine learning uses three classes of techniques: neural networks, symbolic learning, and genetic algorithms (Chen, 1995). Data mining benefits from these methods but differs from the objective pursued: extracting patterns, describing trends, and predicting behavior. Data mining techniques are applied in a wide range of information based on subject matter wherein exist huge amounts of data. In this sense, data mining techniques used in the World Wide Web are called web mining; used in text are called text mining, and used in libraries are called bibliomining.

The application of bibliomining tools is an emerging trend that can be used to understand patterns of behavior among library users and staff, and patterns of information resource used throughout the library (Nicholson & Stanton, 2006). Bibliomining is highly recommended in providing useful and necessary information for library management requirements, focusing on the professional librarianship issues, but is highly database-technical dependent (Shieh, 2010). Bibliomining can also be used to provide a comprehensive overview of the library workflow in order to monitor staff performance, determine areas of deficiency, and predict future user requirements (Prakash, Chand, & Gohel, 2004). The resulting information gives the possibility to perform scenario analysis of the library system, where different situations that need to be taken into account during a decision-making process are evaluated (Nicholson, 2006a). An additional application is to standardize structures and reports in order to share data warehouses among groups of libraries, allowing libraries to benchmark their information (Nicholson, 2006a). Therefore, in order to improve the interaction quality between a library and its users, the application of data mining tools in libraries is worth pursuing (Chang & Chen, 2006). In literature, several case studies describe different approaches to analyze digital library collection based on data mining techniques; however, only a few studies are presented in literature which regard these techniques as a support tool for decision-making in services and collection (Decker & Höppner, 2006; Laitinen & Saarti, 2012). In most studies the authors conclude that the difficulties arise on deciding which data sources will be included, as well as on integrating the data coming from different platforms and applications.

1.2.6 Resource allocation and budgeting

Resource allocation, according to Barbara Blake Gonzalez (2011), is simply the most complex process of decision-making. In libraries, William B. Rouse (1975) argues that resource allocation can be performed at several levels. For instance, priorities within services or processes are defined on the lowest level, such as the number of librarians or computers for the reference services. At the intermediate level, the decisions are among the different services or processes. In this stage the concerns are about how to deal with the competition of resources, such as collection versus staff. The highest level relates to the competition between the library and other institutional departments. The objective in each stage of resource allocation is to assign funds in the most effective way in order to accomplish the objectives of the institution (Bookstein, 1974; Rouse, 1975). Additional elements in economic decision-making include strategic planning and budgeting. In turn, Bowen (1971) argues that planning and budgeting should be a closely integrated process.

A budget is an indispensable tool for management when aligning resource allocation with institution's priorities. Unfortunately, budgeting is always a complex process since it has to deal with limited resources and growing requirements (Linn, 2007; Wise & Perushek, 1996). A budgeting process, unlike resource allocation, is usually a "top down" approach, as this is mostly directed by economic conditions and university priorities expressed by both institutional authorities and library administrators. Typical budgeting activities involve planning, control, coordination, communication, and prioritization of the resource allocation. Academic libraries, in particular, struggle to make budget decisions in a time of scarce resources, dealing with budgets that tend to decrease or, in the better case, tend to remain constant. In addition aspects like inflation, new information requirements, and increased cost of materials make the library budgeting process rather difficult (Chan, 2008; Sudarsan, 2006b).

Librarians have been discussing the best approach to allocate funds to their collection for a long time now. As a result, many budgeting system approaches to allocate resources can be found in

literature. In academic libraries, these budgeting systems are often mixed, incorporating two or more budgeting strategies such as formula, program-based, zero-based, incremental line-item, performance based and responsibility center based budgeting (Blake Gonzalez, 2011; Linn, 2007). This combination is applied, for example, because one method can be used externally when requesting for funds and a different method can be used when distributing those funds internally. When considering these various options it is wise to keep in mind Green and Monical's (1985) remark, "There are probably as many different ways of allocating resources in institutions of higher education as there are presidents of these institutions".

1.3 Aims of the Project

This section provides an overview of the purpose of the PhD study, followed by posing research questions related to the topic of the study.

1.3.1 Purpose Statement

E-content revolution, technological advances, and ever-shrinking budgets oblige libraries to efficiently allocate their limited resources among collection and services. Unfortunately, this resource allocation is a complex process due to the diversity of constraints, data sources, and formats requiring analysis prior to decision-making, as well as the lack of efficient methods of integration.

The main purpose of this study is to develop an integrated model that can support libraries in making optimal budgeting and resource allocation decisions among their services and collection through a holistic analysis. To meet this goal, a holistic structure and the required toolset to holistically assess academic libraries is firstly proposed to collect and organize the data from an economic point of view. To do so, a four-pronged theoretical framework is used in which the library system and collection are analyzed from the perspective of users and internal stakeholders. Secondly, a data warehousing approach is proposed to integrate, process, and store the holistic-based collected data. Ultimately, several techniques to visualize and analyze the stored data that can help libraries in their decision-making, such as reporting and using data mining tools and optimization models, are explored and implemented. By proposing a holistic approach, this research study aims to provide an integrated solution that assists library managers to make economic decisions based on an "as realistic as possible" perspective of the library situation.

1.3.2 Research Questions

The four major research questions (RQ) addressed in this dissertation are the following:

RQ1: How to collect data in a structured manner, covering the key aspects of a library, and at the same time, facilitating the understanding and replication of the data collection process.

In literature, many resource allocation approaches have been proposed; however, most of them mainly focus on economic allocation for either physical or digital collections separately. The answer to this question is addressed in Chapter 2, where a holistic architecture for data collection is proposed based on an extensive literature review. This holistic approach assesses the library collection and services by analyzing key elements including service performance analysis, quality control, collection usage analysis, and information retrieval effectiveness. One of the advantages of this holistic assessment architecture is the ease of understanding, completeness, replicability, and applicability to both physical and digital resources. An example of organizing and collecting the information based on this holistic approach is presented in Appendix A.

RQ2: How to calculate the cost of library services based on a formal costing analysis in a way that can be widely and effectively applied, while minimizing the required resources.

Libraries have a long history in collecting statistics and data, extensive enough for filling all the quadrants of the holistic assessment structure: service, collection, quality, and usage analysis. However, this research study not only requires including these data but also the cost of services and collection as the basis for an optimal resource allocation and budgeting process. For some aspects like library collection, the cost is normally the same no matter how often the collection is accessed because of a fixed subscription and purchasing cost; but for others like library services, it presents a great challenge. Libraries are not used to performing formal costing analyses to their services and processes. This research question is first analyzed in Chapter 3 with a comprehensive state-of-the-art account of costing systems with special emphasis on TDABC. By describing the TDABC implementation in several library processes, Chapters 4 and Appendix B present TDABC as a useful costing system for librarians and library managers who want to perform a cost analysis in a formal and accurate manner, while keeping simplicity and ease of implementation. Eventually, Chapter 5 addresses RQ2 by utilizing TDABC to maximize process benchmarking.

RQ3: What architecture is adequate to store the data collected through the holistic approach from different sources and formats that enables to analyze and maintain big quantities of data?

With a complete framework for data collection, these data coming from multiple sources, and therefore with different formats, need to be integrated and stored in an adequate solution for decision support. Subsequently, such solution should allow data manipulation, analysis, and visualization. Unfortunately, such integration presents a big challenge to be addressed since these different data sources normally use dissimilar formats and access methods. To overcome these shortcomings, a data warehousing approach is proposed in Chapter 6 to integrate, filter and process all the information extracted from many different systems based on a holistic approach. In this chapter, the design of an integrated decision support system based on data warehousing techniques is presented through a case study.

RQ4: What tools and strategies can be used to visualize and analyze strategic information to support libraries in decision-making?

Strategic data stored in the data warehouse are used for different purposes: 1) data visualization and reporting, allowing library managers to publish library indicators in a simple and quick manner by using online reporting tools; 2) sophisticated data analysis through the use of data mining tools; and 3) input for optimization models. An implementation of data visualization and reporting is indicated in Chapter 6. Data mining or bibliomining techniques analyze large information databases and discover implicit, but potentially, useful information. Data mining has the capability to uncover hidden relationships and to reveal unknown patterns and trends by digging into large amounts of data. Chapter 7 addresses RQ4 by reporting the results of an exploratory and systematic literature review of the use of data mining applications in academic libraries. Based on these results and implications, an experimental use of bibliomining techniques to support library decision-making is described in Chapter 8. Eventually, preliminary experiences of an optimization model implementation to solve materials budget allocation are reported in Chapter 9.

1.4 Case Studies

Most studies about resource allocation in libraries have primarily been performed in Europe and North-America. Up to now, no Ecuadorian or Latin-American studies have been documented in this

area. For decades, public Ecuadorian universities have suffered financial limitations. In addition to this difficult situation, 85% of those limited resources have been taken up by salaries¹, displacing the library funds allocation to a “second fiddle”. Knowing that one of the determining facts for the development of science and technology is the possibility to have access to scientific knowledge, libraries become key players for their access and dissemination. Therefore, the proper allocation of limited resources in Ecuadorian libraries becomes a very important process that needs to get the attention that is deserved. The following paragraphs present an overview of the selected case studies, e.g., two libraries were selected as main case studies: one of the biggest academic libraries in Ecuador and a big, modern academic library in Belgium.

The University of Cuenca (UC) library, or Regional Documentation Centre “Juan Bautista Vazquez” (or CDRJBV for its acronym in Spanish), is considered one of the most modern and biggest libraries in Ecuador. Its collection, principally supporting the University teaching mission, consists of about 250,000 books (i.e., 18 titles per student which is far above the national ratio), digital databases, and multimedia contents. CDRJBV, daily visited by an average of 1,200 students, is operated by 20 full-time staff members distributed in the main library and two branches².

In general, at the University of Cuenca, funds for library collection development are allocated by faculties; each faculty decides what to subscribe/acquire and what to unsubscribe, generally following historical spending patterns, and in some cases, based on their own finances and priorities. CDRJBV does not manage its own budget; only a small percentage of funds are assigned to the library for operative expenses. In this way, librarians, in many occasions, have failed to answer the question of why more resources are allocated to one faculty collection than another. This historical spending pattern favors certain faculties and knowledge areas and negatively affects newer and emerging faculties or departments within the University.

On the other hand, CDRJBV has come to a crucial stage within its development. Thanks to the Cooperation Program VLIR-IUC (Council of Flemish Universities - Institutional University Cooperation) since 2008, the library is working on the improvement of its services, including new technological infrastructure, organizational structure and information systems. In addition, the current Ecuadorian government has introduced an aggressive strategy to improve quality education, making universities pay attention to libraries for accreditation purposes. Despite these favorable opportunities, there is still little progress in terms of library development. Therefore, under the current functioning conditions and limited offered services in CDRJBV, a stand-alone case study is insufficient to validate the resource allocation model. It is necessary to work in parallel with libraries that already handle a broader range of services. Therefore, these other libraries were chosen as additional case studies to analyze and validate each stage of the research study. One of the libraries selected for this purpose was the Arenberg Campus Library of the KU Leuven.

The Arenberg Campus Library (Campusbibliotheek Arenberg - CBA) is considered as one of the largest and most modern libraries into the areas of science and engineering in Belgium. The library has a collection of one million books, reference works, and additionally offers electronic and multimedia facilities. The CBA staff, approximately 19 full-time equivalent employees (FTE) provide service to about 10,000 potential customers (Bogaerts, Dekeyser, & Holans, 2002). CBA provides its services to the faculties of science, engineering, bioscience engineering, kinesiology, and rehabilitation sciences. To improve cost efficiency and effectiveness, CBA has been forced to find new strategies to deliver its services, such as the use of new technologies, improving access to e-journals and databases, automation of repetitive processes and deployment of new digital and physical services. However, library budget cuts urge the CBA management to keep improving its understanding and selection of the information collected for budget decision-making.

¹ Interview with CFO of the University of Cuenca

² Interview with Director of the Library of the University of Cuenca

To deeply analyze the cost of library services, the processes of the Biomedical Library of the University of Ghent (Kenniscentrum voor de Gezondheidszorg Gent - KCGG) are also incorporated in the study in order to corroborate the value of the proposed costing technique. KCGG³ is the faculty library for the Faculties of Medicine, Health, and Pharmaceutical Sciences. Founded in 1991 as a joint initiative of the University of Ghent and the Ghent University Hospital (Universiteit Ziekenhuis - UZ), KCGG also holds the collection of the UZ campus departmental libraries. Besides these two libraries in Belgium, used to develop the costing model, and the CDRJBV in Ecuador, utilized to validate the initial model, ten additional Ecuadorian academic libraries were also analyzed through the proposed costing technique in order to improve the model. These universities were selected at the beginning of this research study. The selected universities were classified as Category A according to CONEA's evaluation (National Council for Accreditation of Higher Education - Ecuador) (Consejo Nacional de Evaluación y Acreditación de la Educación Superior del Ecuador, 2009).

1.5 Dissertation Overview

This section provides an overview of the dissertation structure. Although this PhD dissertation is a compilation of articles, and consequently all chapters can be read independently, there is a logical order, depicted in Figure 1.3 in which the investigation was developed and that is reflected in this dissertation. The dissertation consists of eleven chapters. The Introduction and the Conclusion sections are Chapters 1 and 10. Chapter 2 meets RQ1 by detailing a holistic structure and the required set of tools to holistically assess academic libraries from an economic perspective; Chapters 3, 4 and 5 address RQ2 and complement the holistic structure by analyzing and implementing a costing system which is fundamental to calculate both, the costs of processes and library services. Chapter 6 meets RQ3 by proposing and implementing a data warehousing architecture to integrate, process, and store the holistic-based collected data. The forth research question is addressed in Chapters 7, 8 and 9 by describing and implementing several techniques to analyze and visualize the compiled data.

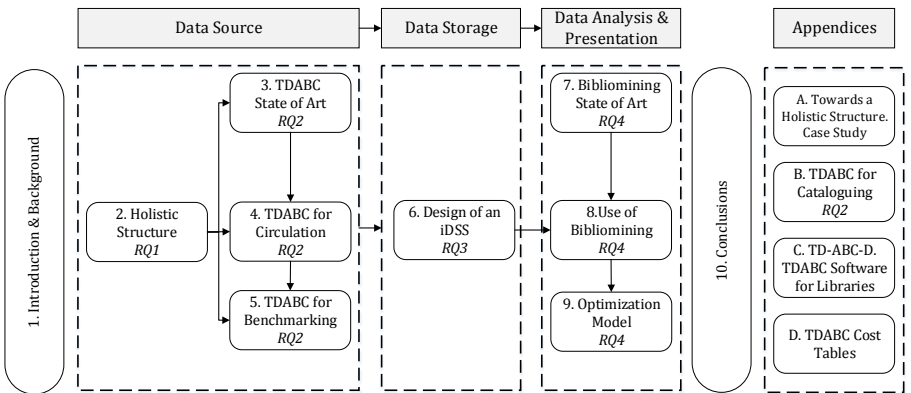


Figure 1.3: Structure of the dissertation

The second part of this dissertation starts in Chapter 2 by detailing an evaluation framework approach and the required set of tools to holistically assess academic libraries from an economic point of view. Additionally, a data warehouse architecture is proposed to integrate process and store the holistic-based collected data.

³ Biomedical Library of the University of Ghent - <http://www.kcgg.ugent.be/>

In Chapter 3, a comprehensive literature review of Time-Driven Activity Based Costing (TDABC) is provided. TDABC is a relatively new tool to improve the cost allocation to products and services. After a brief overview of traditional costing and activity based costing systems (ABC), a detailed description of the TDABC model is given and a comparison made between this methodology and its predecessor ABC. Thirty-six empirical contributions using TDABC over the period 2004-2012 are analyzed. Findings are grouped according to the main areas of application of the method such as logistics, manufacturing, services, health, hospitality and non-profit services. Potential benefits and challenges are identified.

Chapter 4 presents a detailed case study and corresponding analysis conducted using TDABC on the lending and returning processes of an academic library in Belgium. The applicability of TDABC in academic libraries is illustrated with special attention to large-scale libraries. To do so, the chapter firstly provides a theoretical background of TDABC, including its main characteristics and limitations. Then, the TDABC implementation in the case study is analyzed, identifying key benefits and deployment limitations faced during the process. Finally, some conclusions and recommendations for future work are also provided.

Chapter 5 provides more detailed insight of using TDABC to benchmark library processes. The chapter starts by explaining the different steps involved in implementing TDABC in academic libraries, followed by a comparison of the work processes and procedures. Next, the benefits and detriments encountered when implementing the TDABC model in two medium-sized academic libraries in Belgium are described. For this purpose, four main library functions are studied: acquisition, cataloging, circulation, and document delivery. This chapter ends by offering a number of recommendations where process flow improvements were discovered.

The third part of this dissertation starts in Chapter 6 with the analysis and design of an integrated decision-support system based on data warehouse techniques for an academic library. To this end, the holistic approach described in Chapter 2 is used for data collection. Based on this proposed approach, a set of queries of interest is described to be issued against the integrated system. Then, relevant data sources, formats and connectivity requirements for a particular case study are identified. Next, a data warehouse architecture is proposed to integrate, process, and store the collected data transparently. Eventually, the stored data are analyzed through reporting techniques, specifically on-line analytical processing tools.

The fourth part of this dissertation starts in Chapter 7. This chapter introduces data mining as an additional technique to analyze and visualize strategic library information. A comprehensive literature review and classification scheme for data mining techniques applied to libraries is provided. Forty-one empirical contributions over the period 1998-2014 are analyzed for their direct relevance. To do so, a detailed explanation of the research methodology adopted is first provided. This is followed by a description of the proposed method for classifying data mining applications in libraries. Classification results are then presented and discussed. The chapter finalizes by presenting limitations of the study and by outlining research implications and prospects for future research developments.

Chapter 8 reports an experimental use of data mining techniques for library decision-making. The three-layer data warehouse architecture described in Chapter 6 is used to integrate, process, and store the collected data. First, the theoretical background and related work is briefly presented. Next, the chapter describes the selection of methodology for the construction of the DW, as well as the architecture and implementation framework. The stored data are then queried and analyzed utilizing three data mining techniques: regression, clustering, and classification. Chapter 8 finalizes by summarizing lessons learned and identifying future challenges and directions.

Eventually, Chapter 9 documents early experiences of developing an optimal resource allocation model to distribute resources among the different processes of a library system. Specifically, this chapter addresses the problem of allocating funds for digital collection among divisions of an academic library. The chapter defines an optimization model for the problem with an objective of

maximizing the usage of digital collection over all library divisions subject to a single collections budget. An application of this model to an academic library in Belgium is discussed.

References

- ACRL Research Planning and Review Committee. (2012). 2012 top ten trends in academic libraries: A review of the trends and issues affecting academic libraries in higher education. *College & Research Libraries News*, 73(6), 311–320.
- ACRL Research Planning and Review Committee. (2013). *Environmental Scan 2013* (pp. 1–45).
- ACRL Research Planning and Review Committee. (2014). Top trends in academic libraries: A review of the trends and issues affecting academic libraries in higher education. *College & Research Libraries News*, 75(6), 294–302.
- Agee, J. (2005). Collection evaluation: A foundation for collection development. *Collection Building*, 24(3), 92–95. <http://doi.org/10.1108/01604950510608267>
- Allen Press, Inc. (2012). 2012 Study of Subscription Prices for Scholarly Society Journals: Society Journal Pricing Trends and Industry Overview (p. 20 p.). Allen Press, Inc. Retrieved from http://allenpress.com/system/files/pdfs/library/2012_AP_JPS.pdf
- Alvite, L., & Barrionuevo, L. (2010). *Libraries for Users: Services in Academic Libraries*. Elsevier.
- Bertot, J. C. (2011). Concluding Comments: 2010 Library Assessment Conference. *The Library Quarterly*, 81(1), 127–128. <http://doi.org/10.1086/657502>
- Blake Gonzalez, B. (2011). Resource Allocation Strategies in Doctoral/Research University (extensive) Libraries. THE GEORGE WASHINGTON UNIVERSITY. Retrieved from <http://gradworks.umi.com/34/33/3433988.html>
- Bogaerts, W., Dekeyser, R., & Holans, L. (2002). Logistics, physical organization of a library, problems related to new buildings: The CBA case - "Campusbibliotheek Arenberg." In *Proceedings of the International Symposium "Science & Engineering Libraries for the 21st century"* (pp. 97–112). Leuven, Belgium: Leuven University Press.
- Bookstein, A. (1974). Allocation of resources in an information system. *Journal of the American Society for Information Science*, 25(1), 52–58. <http://doi.org/10.1002/asi.4630250107>
- Borgman, C. L. (2003). The invisible library: Paradox of the global information infrastructure. *Library Trends*, 51(4), 652–674.
- Bowen, W. G. (1971). The Role of the Business Officer in Managing Educational Resources. *The Management Challenge: Now and Tomorrow. Managing Educational Programs*. NACUBO Professional File, 2(3). Retrieved from <http://eric.ed.gov/?id=ED113993>
- Brook, J., & Salter, A. (2012). E-books and the Use of E-book Readers in Academic Libraries: Results of an Online Survey. *Georgia Library Quarterly*, 49(4). Retrieved from <http://digitalcommons.kennesaw.edu/glq/vol49/iss4/10>
- Byfield, T., Project, C. H., & Stanway, P. (2004). *The Quest for the City: A.D. 740 to 1100 : Pursuing the Next World, They Founded this One*. Christian History Project.
- Casson, L. (2002). *Libraries in the Ancient World*. Yale University Press.
- Chang, C.-C., & Chen, R.-S. (2006). Using data mining technology to solve classification problems: A case study of campus digital library. *Electronic Library*, The, 24(3), 307–321. <http://doi.org/10.1108/02640470610671178>
- Chan, G. R. Y. C. (2008). Aligning collections budget with program priorities: A modified zero-based approach. *Library Collections, Acquisitions, and Technical Services*, 32(1), 46–52. <http://doi.org/10.1016/j.lcats.2008.06.001>
- Chen, H. (1995). Machine learning for information retrieval: Neural networks, symbolic learning, and genetic algorithms. *Journal of the American Society for Information Science*, 46(3), 194–216. [http://doi.org/10.1002/\(SICI\)1097-4571\(199504\)46:3<194::AID-ASIA>3.0.CO;2-S](http://doi.org/10.1002/(SICI)1097-4571(199504)46:3<194::AID-ASIA>3.0.CO;2-S)
- Ching, S. H., Leung, M. W., Fidow, M., & Huang, K. L. (2008). Allocating costs in the business operation of library consortium: The case study of Super e-Book Consortium. *Library Collections, Acquisitions, and Technical Services*, 32(2), 97–103. <http://doi.org/10.1016/j.lcats.2008.08.013>
- Consejo Nacional de Evaluación y Acreditación de la Educación Superior del Ecuador. *Evaluación de Desempeño Institucional de las Universidades y Escuelas Politécnicas del Ecuador*, Pub.

- L. No. Mandato Constituyente No. 14 (2009). Retrieved from http://www.educacionsuperior.gob.ec/wp-content/uploads/downloads/2012/07/Extracto_informe_CONEA.pdf
- Cooper, R., & Kaplan, R. S. (1988). Measure costs right: Make the right decision. *Harvard Business Review*, 66(5), 96–103.
- Curtis, M. B., & Joshi, K. (2011). Developing a Data Warehouse: Some Guidelines and Suggestions. *The Review of Business Information Systems*, 3(2), 13–26.
- Dalci, I., Tanis, V., & Kosan, L. (2010). Customer profitability analysis with time-driven activity-based costing: a case study in a hotel. *International Journal of Contemporary Hospitality Management*, 22(5), 609–637. <http://doi.org/10.1108/09596111011053774>
- Danskin, A. (2007). "Tomorrow never knows": the end of cataloguing? *IFLA Journal*, 33(3), 205–209. <http://doi.org/10.1177/0340035207083301>
- Decker, R., & Höppner, M. (2006). Strategic planning and customer intelligence in academic libraries. *Library Hi Tech*, 24(4), 504–514. <http://doi.org/10.1108/07378830610715374>
- Dejnega, O. (2011). Method Time Driven Activity Based Costing - Literature Review. *Journal of Applied Economic Sciences*, 6(1(15)), 7–15.
- Delaney, G., & Bates, J. (2014). Envisioning the Academic Library: A Reflection on Roles, Relevancy and Relationships. *New Review of Academic Librarianship*, 1–22. <http://doi.org/10.1080/13614533.2014.911194>
- Demeere, N., Stouthuysen, K., & Roodhooft, F. (2009). Time-driven activity-based costing in an outpatient clinic environment: Development, relevance and managerial impact. *Health Policy*, 92(2–3), 296–304. <http://doi.org/10.1016/j.healthpol.2009.05.003>
- Ellis, E. L., Rosenblum, B., Stratton, J. M., & Ames-Stratton, K. (2014). Positioning academic libraries for the future: A process and strategy for organizational transformation. Presented at the 35th IATUL Conference (International Association of Scientific and Technological University Libraries), Espoo, Finland. Retrieved from <http://hdl.handle.net/1808/14141>
- Ellis-Newman, J. (2003). Activity-Based Costing in user services of an academic library. *Library Trends*, 51(3), 333–348.
- Ellis-Newman, J., & Robinson, P. (1998). The cost of library services: Activity-based costing in an Australian academic library. *The Journal of Academic Librarianship*, 24(5), 373–379. [http://doi.org/10.1016/S0099-1333\(98\)90074-X](http://doi.org/10.1016/S0099-1333(98)90074-X)
- Ernst, D. J., & Segall, P. (1995). Information Resources and Institutional Effectiveness: The Need for a Holistic Approach to Planning and Budgeting. *Cause/Effect*, 18(1), 11–13.
- E. Stewart Saunders. (2003). Cost efficiency in ARL academic libraries. *The Bottom Line*, 16(1), 5–14. <http://doi.org/10.1108/08880450310464009>
- Everaert, P., Bruggeman, W., & De Creus, G. (2008). Sanac Inc.: From ABC to time-driven ABC (TDABC) – An instructional case. *Journal of Accounting Education*, 26(3), 118–154. <http://doi.org/10.1016/j.jaccedu.2008.03.001>
- Everaert, P., Bruggeman, W., Sarens, G., Anderson, S. R., & Levant, Y. (2008). Cost modeling in logistics using time-driven ABC: Experiences from a wholesaler. *International Journal of Physical Distribution & Logistics Management*, 38(3), 172–191. <http://doi.org/10.1108/09600030810866977>
- Gerdson, T. (2002). Activity Based Costing as a Performance Tool for Library & Information Technology Services. In *Proceedings of the 4th Northumbria International Conference on Performance Measurement in Libraries Information Services* (Vol. 4, pp. 119–126). Washington, DC, USA: Association of Research Libraries.
- Girija, N., & Srivatsa, S. K. (2006). A research study: Using Data Mining in knowledge base business strategies. *Information Technology Journal*, 5(3), 590–600. <http://doi.org/10.3923/itj.2006.590.600>
- Goddard, A., & Ooi, K. (1998). Activity-Based Costing and central overhead cost allocation in universities: A case study. *Public Money and Management*, 18(3), 31–38. <http://doi.org/10.1111/1467-9302.00124>
- Green, J. L., & Monical, D. G. (1985). Resource allocation in a decentralized environment. *New Directions for Higher Education*, 1985(52), 47–63. <http://doi.org/10.1002/he.36919855206>

- Guarria, C. I. (2009). How using an allocation formula changed funding allocations at Long Island University. *Collection Building*, 28(2), 44–50. <http://doi.org/10.1108/01604950910953071>
- Guarria, C. I., & Wang, Z. (2011). The economic crisis and its effect on libraries. *New Library World*, 112(5/6), 199–214. <http://doi.org/10.1108/03074801111136248>
- Hand, D. J. (1998). Data Mining: Statistics and more? *The American Statistician*, 52(2), 112–118. <http://doi.org/10.1080/00031305.1998.10480549>
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and techniques* (3rd ed.). Elsevier.
- Heaney, M. (2004). Easy as ABC? Activity-based costing in Oxford University Library Services. *Bottom Line: Managing Library Finances*, The, 17(3), 93–97. <http://doi.org/10.1108/08880450410699701>
- Hobohm, H.-C. (2004). *Knowledge Management: Libraries and Librarians Taking Up the Challenge*. Walter de Gruyter.
- Hui, S. C., & Jha, G. (2000). Data mining for customer service support. *Information & Management*, 38(1), 1–13. [http://doi.org/10.1016/S0378-7206\(00\)00051-3](http://doi.org/10.1016/S0378-7206(00)00051-3)
- Inmon, W. H. (2005). *Building the data warehouse* (4 ed.). John Wiley & Sons.
- Kao, S. C., Chang, H. C., & Lin, C. H. (2003). Decision support for the academic library acquisition budget allocation via circulation database mining. *Information Processing & Management*, 39(1), 133–147. [http://doi.org/10.1016/S0306-4573\(02\)00019-5](http://doi.org/10.1016/S0306-4573(02)00019-5)
- Kaplan, R. S., & Anderson, S. R. (2003). Time-Driven Activity-Based Costing. SSRN eLibrary. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=485443
- Kaplan, R. S., & Anderson, S. R. (2004). Time-Driven Activity-Based Costing - Tool Kit. *Harvard Business Review*, (82), 131–138.
- Kaplan, R. S., & Anderson, S. R. (2007a). The innovation of time-driven activity-based costing. *Journal of Cost Management*, 21(2), 5–15.
- Kaplan, R. S., & Anderson, S. R. (2007b). *Time-Driven Activity-Based Costing: A simpler and more powerful path to higher profits*. Boston, MA, USA: Harvard Business School Press.
- Kennan, M. A., & Wilson, C. (2006). Institutional repositories: review and an information systems perspective. *Library Management*, 27(4/5), 236–248. <http://doi.org/10.1108/01435120610668179>
- Kimball, R. (2006). *The data warehouse toolkit*. John Wiley & Sons.
- Kononenko, I., & Kukar, M. (2007). *Machine Learning and Data Mining*. Elsevier.
- Lai, M.-C., Wang, W.-K., Huang, H.-C., & Kao, M.-C. (2011). Linking the benchmarking tool to a knowledge-based system for performance improvement. *Expert Systems with Applications*, 38(8), 10579–10586. <http://doi.org/10.1016/j.eswa.2011.02.101>
- Laitinen, M., & Saarti, J. (2012). A model for a library-management toolbox: Data warehousing as a tool for filtering and analyzing statistical information from multiple sources. *Library Management*, 33(4/5), 253–260. <http://doi.org/10.1108/01435121211242290>
- Linn, M. (2007). Budget systems used in allocating resources to libraries. *Bottom Line: Managing Library Finances*, The, 20(1), 20–29. <http://doi.org/10.1108/08880450710747425>
- Lynch, C. (2000). From automation to transformation: Forty years of libraries and information technology in higher education. *Educause Review*, 60–68.
- Matthews, J. R. (2011). Assessing Organizational Effectiveness: The Role of Performance Measures. *The Library Quarterly*, 81(1), 83–110. <http://doi.org/10.1086/657447>
- McKendrick, J. (2011). *Funding and priorities: The library resource guide benchmark study on 2011 library spending plans* (p. 40). Chatham, NJ, USA: Unisphere Research, a division of Information Today, Inc. Retrieved from <http://conan.lib.muohio.edu/ebooks/Fundingandpriorities.pdf>
- Mugridge, A. (2012). The Books and Especially the Parchments: Libraries in Society, early Christianity and today. *The ANZTLA EJournal*, 0(9), 134 – 138.
- Mukherjee, A. K. (1966). *Librarianship: Its Philosophy and History*. Asia Publishing House.
- Nicholas, D., Huntington, P., Jamali, H. R., Rowlands, I. H., & Fieldhouse, M. (2009). Student digital information-seeking behaviour in context. *Journal of Documentation*, 65(1), 106–132. <http://doi.org/10.1108/00220410910926149>
- Nicholson, S. (2003). The bibliomining process: Data Warehousing and Data Mining for library decision-making. *Information Technology and Libraries*, 22(4), 146–151.

- Nicholson, S. (2004). A conceptual framework for the holistic measurement and cumulative evaluation of library services. *Journal of Documentation*, 60(2), 164–182. <http://doi.org/10.1108/00220410410522043>
- Nicholson, S. (2006a). Approaching librarianship from the data: Using bibliomining for evidence-based librarianship. *Library Hi Tech*, 24(3), 369–375. <http://doi.org/10.1108/07378830610692136>
- Nicholson, S. (2006b). The basis for bibliomining: Frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services. *Information Processing & Management*, 42(3), 785–804. <http://doi.org/10.1016/j.ipm.2005.05.008>
- Nicholson, S., & Stanton, J. (2006). Bibliomining for library decision-making. In *Encyclopedia of Data Warehousing and Mining* (Second Edition, pp. 100–105). Retrieved from <http://www.igi-global.com/chapter/encyclopedia-data-warehousing-mining/10591>
- Nicholson, S., & Stanton, J. (2009). Bibliomining for library decision-making. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (Second Edition, pp. 341–345). IGI Global. Retrieved from <http://www.igi-global.com/chapter/encyclopedia-information-science-technology-second/13596>
- Niu, X., Hemminger, B. M., Lown, C., Adams, S., Brown, C., Level, A., ... Cataldo, T. (2010). National study of information seeking behavior of academic researchers in the United States. *Journal of the American Society for Information Science and Technology*, 61(5), 869–890. <http://doi.org/10.1002/asi.21307>
- Novak, D. D., Paulos, A., & Clair, G. S. (2011). Data-driven budget reductions: A case study. *The Bottom Line: Managing Library Finances*, 24(1), 24–34. <http://doi.org/10.1108/08880451111142015>
- Pernot, E., Roodhooft, F., & Van den Abbeele, A. (2007). Time-Driven Activity-Based Costing for inter-library services: A case study in a university. *The Journal of Academic Librarianship*, 33(5), 551–560. <http://doi.org/10.1016/j.jalib.2007.06.001>
- Poll, R. (2001). Performance measures for library networked services and resources. *Electronic Library*, The, 19(5), 307–315. <http://doi.org/10.1108/02640470110408887>
- Prakash, K., Chand, P., & Gohel, U. (2004). Application of Data Mining in library and information services (pp. 168–177). Presented at the 2nd Convention PLANNER, Manipur Uni., Imphal: INFLIBNET Centre, Ahmedabad. Retrieved from <http://shodhganga.inflibnet.ac.in/dxml/handle/1944/435>
- Radulescu, L. C. (2003). Managing a Data Warehouse Environment. *Annals. Computer Science Series*, 1(2), 82–93.
- Raju, J. (2014). Knowledge and skills for the digital era academic library. *The Journal of Academic Librarianship*, 40(2), 163–170. <http://doi.org/10.1016/j.jalib.2014.02.007>
- Ross, L., & Sennyye, P. (2008). The library is dead, long live the library! The practice of academic librarianship and the digital revolution. *The Journal of Academic Librarianship*, 34(2), 145–152. <http://doi.org/10.1016/j.jalib.2007.12.006>
- Rouse, W. B. (1975). Optimal resource allocation in library systems. *Journal of the American Society for Information Science*, 26(3), 157–165. <http://doi.org/10.1002/asi.4630260304>
- Setton, K. M. (1960). From Medieval to Modern Library. *Proceedings of the American Philosophical Society*, 104(4), 371–390.
- Shieh, J.-C. (2010). The integration system for librarians' bibliomining. *Electronic Library*, The, 28(5), 709–721. <http://doi.org/10.1108/02640471011081988>
- Singh, H. S. (1998). *Data Warehousing: Concepts, Technologies, Implementations, and Management*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.
- Skilbeck, M., & Connell, H. (2001). Activity based costing: a study to develop a costing methodology for Library and Information Technology activities for the Australian higher education sector (p. 124 p.). Information and Education Services Division, University of Newcastle. Retrieved from <http://www.dest.gov.au/NR/rdonlyres/49D61919-B87F-4369-9AA3-4DE5876FB32B/4653/libraries.pdf>
- Stephens, A. K. (1998). Introduction. *The Acquisitions Librarian*, 10(20), 1–3.
- Stockwell, F. (2000). *A History of Information Storage and Retrieval*. McFarland.
- Stouthuysen, K., Swiggers, M., Reheul, A.-M., & Roodhooft, F. (2010). Time-Driven Activity-Based Costing for a library acquisition process: A case study in a Belgian University. *Library*

- Collections, Acquisitions, and Technical Services, 34(2-3), 83-91.
<http://doi.org/16/j.lcats.2010.05.003>
- Sudarsan, P. K. (2006a). A resource allocation model for university libraries in India. *Bottom Line: Managing Library Finances, The*, 19(3), 103-110.
<http://doi.org/10.1108/08880450610682527>
- Sudarsan, P. K. (2006b). A resource allocation model for university libraries in India. *Bottom Line: Managing Library Finances, The*, 19(3), 103-110.
<http://doi.org/10.1108/08880450610682527>
- Sumathi, S., & Sivanandam, S. N. (2006). *Introduction to Data Mining and its applications*. Springer.
- Townley, C. T. (2001). *Knowledge Management and Academic Libraries*. *College & Research Libraries*, 62(1), 44-55.
- Trumble, K. (2003). *The Library of Alexandria*. Houghton Mifflin Harcourt.
- Tse, M. S. C., & Gong, M. Z. (2009). Recognition of idle resources in time-driven activity-based costing and resource consumption accounting models. *Journal of Applied Management Accounting Research*, 7(2), 41-54.
- Uhlendorf, B. A. (1932). *The Invention of Printing and Its Spread till 1470: With Special Reference to Social and Economic Factors*. *The Library Quarterly: Information, Community, Policy*, 2(3), 179-231.
- Walters, W. H. (2013). E-books in Academic Libraries: Challenges for Acquisition and Collection Management. *Portal : Libraries and the Academy*, 13(2), 187.
- Wegmann, G., & Nozile, S. (2009). The Activity-Based Costing method developments: State-of-the art and case study. *The IUP Journal of Accounting Research and Audit Practices*, 8(1), 7-22. <http://doi.org/oai:halshs.archives-ouvertes.fr:halshs-00482137>
- Wise, K., & Perushek, D. E. (1996). Linear goal programming for academic library acquisitions allocations. *Library Acquisitions: Practice & Theory*, 20(3), 311-327.
[http://doi.org/10.1016/0364-6408\(96\)00065-8](http://doi.org/10.1016/0364-6408(96)00065-8)
- Wrembel, R., & Koncilia, C. (2007). *Data Warehouses and OLAP: Concepts, Architectures, and Solutions*. Idea Group Inc. (IGI).
- Ying Wah, T., Hooi Peng, N., & Sue Hok, C. (2007). *Building Data Warehouse*. In *Proceedings of the 24th South East Asia Regional Computer Conference*. Bangkok, Thailand. Retrieved from http://www.ijcim.au.edu/SpecialEditions/v15nSP4/P05SEARCC_BuildingDataWarehous e.pdf
- Zhang, Y. (2010). Developing a holistic model for digital library evaluation. *Journal of the American Society for Information Science and Technology*, 61(1), 88-110.
<http://doi.org/10.1002/asi.21220>

PART II

Data Source

Chapter 2: **Holistic Structure Approach for Data Collection**

Siguenza-Guzman, L., Abbeele, A. V. den, Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2015). A Holistic Approach to Supporting Academic Libraries in Resource Allocation Processes. The Library Quarterly: Information, Community, Policy, 85(3), 295–318. <http://doi.org/10.1086/681612>

This chapter details an evaluation framework approach (Research Question 1) and the required set of tools to holistically assess academic libraries from an economic point of view. Besides a data warehouse architecture is proposed to integrate process and store the holistic-based collected data.

Apart from typographical adjustments, the content of this chapter is identical to the content of the published paper quoted above; where necessary, additional information or remarks are added in footnotes. The layout is adapted for consistency throughout this dissertation. Some redundancy with other chapters is unavoidable as an academic article needs its own introductory sections. This, however, entails the advantage that the chapter can be read separately.

Abstract

E-content revolution, technological advances, and ever-shrinking budgets oblige libraries to efficiently allocate their limited resources between collection and services. Unfortunately, resource allocation is a complex process due to the diversity of data sources and formats required to be analyzed prior to decision-making as well as the lack of efficient methods of integration. The contribution of this article is twofold. We first propose an evaluation framework to holistically assess academic libraries. To do so, a four-pronged theoretical framework is used in which the library system and collection are analyzed from the perspective of users and internal stakeholders. Second we present a data warehouse architecture that integrates, processes, and stores the holistically based collected data. By proposing this holistic approach, we aim to provide an integrated solution that assists library managers to make economic decisions based on a perspective of the library situation that is as realistic as possible.

Contributions of the first author

The first author's contributions are: the literature study on holistic evaluation, the new holistic approach proposed, set of tools and methodologies to evaluate each quadrant, data warehouse architecture, and conclusions.

2.1 Introduction

Amid limited funding resources, libraries strive to efficiently deal with technological advances and e-content revolution (Bertot, 2011). In fact, academic libraries face hard budget constraints due to the global economic crisis (Sudarsan, 2006a; McKendrick, 2011). This dilemma stems from library services usually being “free of charge” but not free of costs and strongly dependent on public funding (Stouthuysen et al., 2010). As a result, despite cuts, mergers, and budget freezes, libraries must create, maintain, and improve their services (Cottrell, 2012; Cox, 2010; Guarria & Wang, 2011). Furthermore, the latest technological advances and e-content revolution, such as the growing presence of e-books, and the proliferation of tablets and mobile devices, have influenced the manner in which information is disseminated and consumed (Allen Press, Inc., 2012; Brook & Salter, 2012). As a consequence, academic libraries are rapidly reallocating budgets from print to digital resources. For example, David Nicholas, Ian Rowlands, Michael Jubb, and Hamid R. Jamali (2010), report that although e-books still account for a small proportion of total spending - approximately 5%- this figure is rising rapidly. Online content facilitates managing information, is often cost effective, and more easily accessible than printed resources; however, it also contributes to increasing the complexity of the resource allocation process (Chan, 2008; Guarria, 2009; Poll, 2001). For instance, one problem with a subscription-based digital library collection is the variability of yearly prices that has evolved in the last years (Allen Press, Inc., 2012). Furthermore, in order to provide these e-services, academic libraries have to deal with challenges such as the lack of uniformity in license terms, lease conditions, access restrictions, and librarians’ expectations (Walters, 2013).

Dynamic components such as the e-content revolution, technological advances, and ever-shrinking budgets constantly force libraries to be more innovative in providing, justifying, and evaluating the effectiveness of their services (ACRL Research Planning and Review Committee, 2010; Blixrud, 2003). David J. Ernst and Peter Segall (1995) state that institutions in these difficult circumstances are called upon to develop a strategic and well-coordinated budget plan by means of a “holistic approach”. The objective of the holistic approach is to help organizations define a set of measures that reflect their objectives and assess their performance appropriately (Matthews, 2011). The holistic approach requires interconnecting all necessary components in a way that responds to both shrinking resources and dynamic library services. Unfortunately, interconnecting and analyzing all the heterogeneous data sets are complex processes due to the large number of data sources and volume of data to be considered. Therefore, the aim of the paper is twofold. First, we present a holistic structure and the required set of tools for collecting data from an economic point of view. The holistic structure uses a theoretical framework based on a two-dimensional evaluation matrix (Figure 2.1) in which the library system and its collection are analyzed from an internal and external perspective. Secondly, we propose the design of an integrated decision support system that integrates, processes, and stores the collected data.

		Topic	
		Library System	Use
Perspective	Internal (Library System)	1. What does the library system consist of?	4. How is the library system manipulated?
	External (Users)	2. How effective is the library system?	3. How useful is the library system?

Figure 2.1: Conceptual matrix for holistic measurement (Nicholson, 2004)

2.2 Theoretical background

A budget is a financial plan that normally reflects the organization's priorities; through this, managers boost important activities by allocating enough resources to them and ration resources for less-important areas of an organization (Linn, 2007). Many approaches of budgeting systems have been proposed in literature, such as incremental-line-item, formula-based, mathematical decision model-based, zero-based, and "home-made" resource allocation methods (Linn, 2007; Smith, 2008). Each budgeting system functions differently and is often used in combination with other methods. For instance, one method can be used externally when applying for funds, and another can be used when distributing those funds internally.

In the case of academic libraries, collection budgets used to be allocated by taking into account several factors such as the number of students, circulation of materials, interlibrary loans, number of researchers, and average cost of materials per discipline (Kao et al., 2003). Unfortunately, these indicators to quantify the collection requirements or the usage statistics are not enough anymore. Libraries nowadays must be able to show, on the one hand, their investments and the availability of their resources in producing better results in research and education, and, on the other hand, their effectiveness in delivering library services (Laitinen & Saarti, 2012). To do so, library managers must have enough data to ensure the integration of different areas involved in the library system in order to evaluate and decide how to allocate and prioritize resources to each service or material that a library requires. In this respect, a holistic evaluation to obtain a thorough knowledge of the library system becomes an interesting alternative to be used as a manner in which to organize the data collected for a resource allocation process.

Holism is a concept which emphasizes the importance of the whole and the interdependence of its parts (The American Heritage Dictionaries, 2011). This means that systems work as a whole and cannot be fully understood by analyzing their components separately. If this concept is translated to libraries, holism can be seen as an analysis that emphasizes the importance of the entire library and the interdependence of its processes, collection, and services. Many resource allocation approaches, based on holistic evaluations, have been proposed; however, the majority focuses separately on the economic allocation for physical or digital collections. For instance, F. Wilfred Lancaster (1977, 1988) establishes evaluation procedures only for traditional library services, and Ying Zhang (2010) and Norbert Fuhr, Giannis Tsakonas, Trond Aalberg, Maristella Agosti, Preben Hansen, Sarantos Kapidakis, and Claus-Peter Klas (2007) propose a holistic evaluation model for digital library services. In contrast, Scott Nicholson (2004) proposes a theoretical analysis framework to support libraries in gaining a more thorough and holistic understanding of their users and services for both digital and physical services. As can be seen in Figure 2.1, Nicholson proposes an evaluation matrix with four quadrants in which columns represent the topics library system and collection, and rows represent the perspectives of library staff and users. Because of the ease of understanding, completeness, and applicability to both physical and digital resources, this theoretical framework is adopted as a basis to propose a holistic structure for data collection and in turn, uses these data sets as an input for an integrated decision support system.

The following items briefly describe the main features of each quadrant proposed by Nicholson:

1. If the library system is analyzed from an internal perspective, the question to be answered is **"What does the library system consist of?"** This is a traditional type of analysis that can include bibliographic collection aspects, organizational flows, computer interfaces, processes, staff, and resources.
2. The second quadrant evaluates the user's perception about service quality. Aboutness, effectiveness and usability of the library services are the main aspects studied. The question to be answered is **"How effective is the library system?"**
3. The third quadrant is centered on **"How useful is the library system?"** This quadrant allows quantification of the impact of the library collection on its users, thus providing library managers with a better basis for decision making when acquiring new bibliographic

materials. By evaluating the current bibliographic collection, libraries may discover possible gaps and plan future collection development (Agee, 2005).

4. The fourth quadrant aims to answer the question **“How is the library system manipulated?”** This quadrant analyzes the use patterns followed to manipulate the library system. For instance, in digital library services, unlike circulation patterns in traditional services, it is possible to track everything users do within the library system, allowing libraries not only to know what users retrieve but also what they looked for and could not receive.

Thus, by incorporating this simple, but at the same time powerful, theoretical framework to organize the data collection required, this study ensures that evaluating the collection and services in academic libraries is based on a holistic model.

The remainder of this article is divided into three sections. Section 2.3 describes the data collection procedure to holistically analyze academic libraries from an economic perspective. Section 2.4 proposes the design method and structure of a decision support system based on data-warehouse and data-mining technology. Finally, conclusions are drawn in the last final section.

2.3 Data collection through a holistic perspective

In this section, Nicholson’s conceptual matrix is used as a basic reference to propose a structured data collection that ensures a holistic analysis of an academic library from an economic point of view. Based on this structure, a set of tools is provided to collect data for the specific requirements of each quadrant. An example of implementing the proposed holistic approach and tools is presented by Lorena Siguenza-Guzman, Ludo Holans, Alexandra Van den Abbeele, Joos Vandewalle, Henri Verhaaren, and Dirk Cattrysse (2013). The authors highlight the key benefits, challenges, and lessons learned from the implementation of this holistic approach in an academic library in Belgium.

2.3.1 First quadrant: internal perspective of the library system

In this quadrant, the traditional library evaluation (i.e., measurements based on library staff, processes, or systems, but not users) is the main aspect studied. The internal perspective of the library system largely covers the topics related to processes and services carried out within the library system. From an economic perspective, it refers to the need for analyzing the costs incurred and the resources consumed by library processes. Cost analysis techniques, of which the traditional costing system has been one of the most widely used, have been present in libraries for many years.. Jennifer Ellis-Newman, Haji Izan, and Peter Robinson (1996), for instance, describe several studies on library costs that were undertaken in the United States. These studies were carried out with cost allocation models compatible with traditional costing methods. In this type of system, the total cost consists of direct costs such as the cost of consumed resources and direct labor hours, and a percentage of overhead as indirect costs, which are specific costs such as maintenance, marketing, depreciation, training, and electricity. Traditional costing systems are adequate when indirect expenses are low and service variety is limited (Ellis-Newman & Robinson, 1998). However, in environments with a broad range of services such as libraries, indirect costs have increasingly become more important than direct costs (Siguenza-Guzman, Van den Abbeele, et al., 2013).

Seeking to remedy these limitations, libraries started employing more-advanced cost calculation techniques such as Activity-Based Costing (ABC). ABC is an alternative costing system promoted by Robin Cooper and Robert S. Kaplan (1988). Compared with traditional costing methods, ABC performs a more accurate and efficient treatment of indirect costs (Ellis-Newman & Robinson, 1998). In fact, ABC first accumulates overhead costs for each activity and then assigns the costs of the activities to the services causing that activity. An activity for libraries is defined as an event or task undertaken for a specific purpose such as cataloging, loan processing, shelving, and acquisition orders (Ellis-Newman, 2003). An extensive stream of literature describes ABC as a system that

provides interesting advantages for decision making in libraries (Ching, Leung, Fidow, & Huang, 2008; Ellis-Newman, 2003; Ellis-Newman & Robinson, 1998; Gerdson, 2002; Goddard & Ooi, 1998; Heaney, 2004; Novak, Paulos, & Clair, 2011; Skilbeck & Connell, 2001). However, ABC has great limitations, for instance, a high degree of subjectivity in estimating employees' proportion of time spent on each activity; the excessive time, resources, and money for data collection; and the difficulties to model multi-driver activities (Siguenza-Guzman, Van den Abbeele, et al., 2013).

Time-driven activity-based costing (TDABC) is an approach developed by Robert S. Kaplan and Steven R. Anderson (2003) in order to overcome the ABC limitations. TDABC uses only two parameters to assign resource costs directly to the cost objects: 1) the unit cost of supplying resource capacity and 2) an estimated time required to perform an activity (Yilmaz, 2008a). For each activity, costing equations are calculated based on the time required to perform an activity (Yilmaz, 2008a). This time can be readily observed, validated, and then computed by time equations, which are the sum of individual activity times (Kaplan & Anderson, 2007b). By using these equations, all possible combinations of activities can be represented, for example, when different types of services do not necessarily require the same amount of time to be performed. Lorena Siguenza-Guzman, Alexandra Van den Abbeele, Joos Vandewalle, Henri Verhaaren, and Dirk Cattrysse (2013) highlight five TDABC advantages: 1) simplicity in building an accurate model, 2) possibility of using multiple drivers to design cost models for complex operations, 3) good estimation of resource consumption and capacity utilization, 4) versatility and modularity for updating and maintaining the model, and 5) possibility of using the model in a predictive manner.

Up to now, four important studies concerning TDABC in academic libraries have been applied to very specific processes such as the interlibrary loan (Pernot et al., 2007), acquisition (Stouthuysen et al., 2010), circulation (Siguenza-Guzman, Van den Abbeele, Vandewalle, Verhaaren, & Cattrysse, 2014b), and cataloging processes (Siguenza-Guzman, Van Den Abbeele, et al., 2014). In these case studies, TDABC is described as a model that offers a relatively quick and less expensive way to design useful costing models. In addition, Siguenza-Guzman et al. (2013) document the experience of implementing TDABC in 12 library processes. The study highlights three specific advantages: the possibility of disaggregating values per activity, of comparing different scenarios, and of justifying decisions and actions. Two specific challenges are also reported: the significant time required in the data collection, and the staff discomfort with being observed. However, potential solutions to overcome these challenges are also recommended, for instance, the use of a dedicated software tool to perform TDABC analyses, as well as the need for an appropriate communication strategy among library managers and staff to clearly explain the purpose of measurement. In all case studies, the authors conclude that TDABC is, so far, the best system to evaluate costs, processes, and services in academic libraries. TDABC provides accurate information of the library activities, which may help managers get a better understanding of how the library uses its time, costs, and resources. Nevertheless, this information is not sufficient for making management decisions in the library. For instance, consider the following scenario. A library manager is asked to reduce staff due to the high costs of salaries. He consults the costing system, and after a "what-if" analysis, he finds that reference service occupies a surplus of librarians and that by reducing this number, he fulfills the requirement.

Initially, this seems like a good option; however, it provides only a partial solution. The library manager should still consider other aspects, such as the users' perceptions of the service quality falling below the tolerated levels, and the impact of the decision on the entire library system.

2.3.2 Second quadrant: external perspective of the library system

Once the library system has been measured from an internal point of view, the evaluation is balanced by introducing the users' perspectives. By doing this, the framework allows library managers to see beyond the system, staff, or processes and to understand what users really need and desire from the services performed by a library. Nicholson (2004) proposes to evaluate the

aboutness, pertinence, and usability of a library system, including both physical and digital resources. Aboutness refers to analyzing the relevance of library resources and services to their users. It is based on the users' personal judgments of the conceptual relatedness between the users' needs and services offered (Kowalski, 2011). Pertinence takes into account the user and the situation in which the service is to be used. It assumes that users can only make valid judgments about the suitability of services for solving their information needs (Kowalski, 2011). Finally, usability refers to evaluating the library system reliability, meaning whether it can be used without problems.

Libraries have a long history of collecting users' statistics to monitor service quality (Horn & Owen, 2009). In literature, different approaches have emerged (Nitecki & Hernon, 2000); for instance, one approach is centered on the use of SERVQUAL (for SERVICE QUALity) measurements. SERVQUAL is a popular tool from the 1980s developed for assessing service quality in the private sector. This model uses the service quality gap theory proposed by Valarie A. Zeithaml, A. Parasuraman, and Leonard L. Berry (1990) to summarize a set of five gaps showing the discrepancy between perceptions and expectations of customers and managers. Danuta A. Nitecki and Peter Hernon (2000) note that by applying this instrument, libraries gain knowledge about the customer conceptualization of what a service should deliver and how well the service complies with idealized expectations. Another approach is based on the work of Peter Hernon and Ellen Altman (1996, 2010), who build their analysis on an extensive set of expectations around the gaps theory to look at the service nature of libraries. They suggest a pool of more than 100 candidate service attributes from which staff can select a subset potentially having the greatest relevance to their library (Nitecki & Hernon, 2000). An additional approach, described by Joseph R. Matthews (2013), combines data about the library use and library services with other data available on the academic campus. For instance, the author suggests that for university students, library use and services should correlate with either direct or indirect measures of student achievement. Examples of direct measures include the capstone experience, use of a portfolio, or a standardized exam. Indirect measures could include students' grade point averages, success in graduate school exams, and graduate student publications.

Stephanie Wright and Lynda S. White (2007) report the top-five assessment methods used in the past by libraries to measure service quality: statistics gathering, suggestion boxes, web usability testing, user interface usability, and satisfaction surveys. Within these methods, the authors mention that locally designed user satisfaction surveys were widely used; however, they have lately been replaced by surveys developed elsewhere. A detailed description of some of these user-survey methods is provided by Claire Creaser (2006). The author focuses her analysis on the SCONUL user-survey template and the LibQUAL+® surveys. In this article, SCONUL is described as a standard template with a considerable degree of flexibility. SCONUL is offered by the Society of College, and National and University Libraries, and can be adapted to suit local circumstances. LibQUAL+®, likewise, is described as a valuable tool for benchmarking because of its uniformity and limited scope for customization.

The LibQUAL+® survey is a set of services based on web surveys offered by the Association of Research Libraries (ARL). These surveys are based on SERVQUAL measurements, which allow requesting, tracking, understanding, and acting upon users' perceptions of the service quality offered by libraries (Association of Research Libraries, 2012). LibQUAL+®, which was initiated in 2000 as an experimental project, has been applied by more than a thousand libraries around the world, and thanks to its great success it is now considered a standard assessment tool for measuring the quality of services based from users' perceptions (Cook, 2002). This survey helps libraries to assess their strengths and weaknesses, and also benchmark themselves against their peers in order to improve their library services (Franklin, Kyrillidou, & Plum, 2009; Saunders, 2007). The LibQUAL+® survey consists of 22 items or questions that are grouped into three quality dimensions: services provided, physical space, and information resources (Saunders, 2007). The measurement for each perspective uses a scale from 1 to 9. For each question, users give three ratings or levels of service: the minimum expected service quality, the observed or perceived service level, and the desired service level or maximum expectations. Siguenza-Guzman et al. (2013) document the experience of utilizing the LibQUAL+® survey to assess library service

quality. Their study describes the survey results and the action points taken that arose from the survey results. The authors state that although LibQUAL+® provides information on the set of services that require additional attention, some considerations must be taken into account, for example, a data preparation period required to define language and population, granularity to provide benchmarking within branch libraries, and the need of strategies to stimulate participation rates.

By integrating the users' satisfaction criteria with the proposed analysis, library managers now have a broader view of the library system, as they have information about their services and the users' opinions on such services. Assessment methods such as statistics gathering, suggestion boxes, web usability testing, user interface usability, and satisfaction surveys (e.g., LibQUAL+®, or a locally designed survey) are valuable tools to be integrated in our evaluation matrix. The library manager in the aforementioned example may use LibQUAL+®, for instance, to analyze whether the quality of service provided by the reference librarians still lies within the tolerance zone once the changes have been made. Alternatively, libraries can also devise their own instrument, which can be particularly useful for investigating detailed issues (Creaser, 2006). Nevertheless, the selection of the tools to be used in this quadrant depends on their current availability in the library and the decision of library managers whether to include other measures in the model.

2.3.3 Third quadrant: external perspective of the library collection

The goal of this quadrant is to evaluate the usefulness of the library collection. This information allows libraries to gain a more holistic understanding of users' needs and to acquire material that complements current holdings, either improving weak areas or enriching strong collections (Agee, 2005). To do so, two types of measurement are available: 1) through direct contact with the users in order to document which bibliographic materials are valuable to them and 2) through indirect contact by the use of bibliometric analysis (Nicholson, 2004).

Bibliometrics can be defined as the use of mathematical and statistical methods to analyze the use of library information resources. The main focus of bibliometric analyses is on bibliometric distributions of events, such as the productivity of scientific journals, distributions of words in a text, productivity of scientific authors, and circulation of journals within a library or a documentation center (Lafouge & Lainé-Cruzel, 1997). Traditional bibliometrics studies use information about the creation of bibliographic documents such as authors and documents cited, and the metadata associated with them, for example, a general topic area or the specific material in which the metadata appeared. For these studies, the frequency-based analysis is mainly used; nevertheless, many newer bibliometric studies are use visualization techniques and data mining to explore patterns in the creation of these analyses (Nicholson, 2006b).

Of these methods, citation analysis is the best known and most often used, and it is also the one that best meets our requirements for analyzing the use of library information resources. Citation analysis is defined by several authors as: 1) the "wide-ranging area" of bibliometrics that considers the citations to and from documents (Diodato, 1994); 2) a method often used to generate core lists of journals deemed critical to the research needs of an institution (Wallace & Van Fleet, 2001); 3) a technique for counting, tabulating, and ranking the number of times that sources are cited in a document (bibliographies, footnotes, and/or indexing tools) (Edwards, 1999); 4) a method for identifying journals that are often cited, some of which are not from the collection (Feyereisen & Spoiden, 2009). Summarizing the definitions and adjusting them in the context of this research, citation analysis is defined here as a technique for counting, tabulating, and ranking the number of times sources are cited to and from documents in order to analyze the use of a collection. Citation analysis is normally based on samples collected from students' PhD dissertations and master theses. Louise S. Zipp (1996) states that citations from these sources are reliable because they are much more easily and comprehensively gathered and because they reflect the interests of local research groups. Nevertheless, K. Brock Enger (2009, p. 109) recommends caution in the use of

citation analysis. For instance, common lists should be created by comparing library's own results with those of other institutions because students tend to seek only locally owned sources and in many cases may lack the expertise needed to identify the most appropriate sources (Feyereisen & Spoiden, 2009). Likewise, useful information may not be cited, or may be cited by professors, post docs, or researchers in other documents such as syllabi, reports, or books (Feyereisen & Spoiden, 2009), or by those who do not publish, such as undergraduate and graduate students (Duy & Vaughan, 2006). One solution to avoid these omissions is proposed by Robert N. Bland (1980), who suggests citation analysis of the textbooks used in the curriculum.

Vendor-supplied statistics is an additional bibliometric method for evaluating the usefulness of a library collection. The vendor-supplied statistics, also called electronic journal usage data, are usually collected via publisher websites. These lists are normally supplied by vendors as part of their subscription contract. A case study performed by Joanna Duy and Liwen Vaughan (2006, p. 515) advocates the use of this technique to replace the "traditional, expensive and time-consuming manual compilation" of reference lists.

In published journal articles, authors include several references to articles, books, links, and other resources. These citations describe the sources of some concepts or ideas included in the document. At the same time, they help the reader to find relevant information about the topics that were introduced in the original article (He & Cheung Hui, 2002). To measure the value of a journal by the number of citations that a document has had, citation databases have been created. According to Robert A. Buchanan (2006), a citation database serves two purposes: 1) to index the literature using cited articles as index terms, and 2) to measure the number of times a publication has been cited in the literature. A citation database is a warehouse database that analyzes the impact of peer-reviewed literature. The most famous citation databases are Web of Science and Scopus. The selection of a database depends on the research focus. For instance, Scopus covers more relevant journals of medical informatics than does Web of Science (Spreckelsen, Deserno, & Spitzer, 2011).

This study considers that by combining citation analysis, citation database, and vendor-supplied statistics, library administrators will gain an extensive knowledge about the value of their collections. This proposal is also supported by several authors who agree that the use of different methods leads to a more robust indication of collection use and users' needs (Beile, Boote, & Killingsworth, 2004; Duy & Vaughan, 2006; Enger, 2009). The early experiences in developing a project combining these methodologies are documented by Siguenza-Guzman, Holans, et al. (2013). The project analyzes more than 1,200 PhD dissertations submitted over a six-year period (2005-2010). In addition, four databases were created to evaluate citations patterns, publishing patterns, journals downloaded, and journals' impact factors. The authors describe several challenges faced up to now, for instance, 1) the amount of time required to collect the information and incorporate them into databases; 2) the need for a defined standard for naming (e.g., journal's abbreviations); and 3) the need for dedicated software to collect the large amount of information and to evaluate the results.

2.3.4 Fourth quadrant: internal perspective of the library collection

The final quadrant measures users' behavioral aspects within a library system, namely the users' interactions with the system. This interaction is utilized to study users' preferences and to use this information to personalize services (Agostii, Crivellari, & Di Nunzio, 2009).

Transaction log analysis (TLA) is one of the most important and well-known techniques that has been utilized for this purpose. TLA is defined by Thomas A. Peters (1993, p. 42) as "a form of system monitoring and as a way of observing, usually unobtrusively, human behavior." Marcos Gonçalves, Ming Luo, Rao Shen, Mir Ali, and Edward Fox (2002) describe log analysis as a primary source of knowledge in how digital library users actually exploit digital library systems and how systems behave while trying to support users' information-seeking activities. In the context of web search,

the storage and analysis of log files are mainly used to: 1) gain knowledge on users and improve services offered through a web portal without the need to bother users with the explicit collection of information (Agostii et al., 2009), 2) assist users with query suggestions (Kruschwitz et al., 2011), and 3) study the use of online journals and their users' information-seeking behaviors (Jamali, Nicholas, & Huntington, 2005). Measures of usage analysis can include the number and titles of journals used; number of article downloads; usage over time; and a special analysis of subject, date, and method of access (Nicholas, Huntington, Jamali, & Tenopir, 2006).

Many studies have been conducted to corroborate the use of logs analysis to analyze users' behaviors in a digital environment. For instance, Deep Log Analysis (DLA) is a technique employed by Nicholas and colleagues to demonstrate the utility and application of transaction log analysis. The authors conducted a series of studies, such as the comparison of two consumer health sites, NHS Direct Online and SurgeryDoor (Nicholas, Huntington, & Williams, 2002), a comparison of five sources of health information (Nicholas, Huntington, & Homewood, 2003), and a study of the impact of consortia "Big Deals" on users' behaviors (Nicholas, Huntington, & Watkinson, 2005, 2003). Nicholas and colleagues state that web usage logs offer a direct and immediate record of what people have done on a website. Some of the outcomes of DLA include site penetration as the number of items viewed during a particular visit, time online or page view time, type of users identified by IP addresses, academic departments' usage, differentiation among on-campus and off-campus users, and user satisfaction measured by tracking returnees by IP (Nicholas, Huntington, Jamali, & Tenopir, 2006).

Another example on user behavior analysis is presented by Philip M. Davis and Leah R. Solla (2003). The authors report a three-month analysis of usage data for 29 American Chemical Society electronic journals downloaded from Cornell University. They demonstrate that while the majority of users limited themselves to a small number of journals and article downloads, a small minority of heavy users had a large effect on total journal downloads. They conclude that a user population can be estimated by knowing the total use of a journal because of the strong relationship between the number of downloaded articles and the number of users. Nevertheless, the authors use IP addresses as a representation of users, which is not necessarily accurate and might lead to biased results.

Moreover, log analysis can be supported and validated by other types of user studies such as eye-tracking systems to understand users' behaviors in different situations. Eye-tracking systems are devices for measuring eye positions and eye movement (Mehrubeglu, Pham, Le, Muddu, & Ryu, 2011). Hitomi Saito, Hitoshi Terai, Yuka Egusa, Masao Takaku, Makiko Miwa, and Noriko Kando (2009) analyze search behaviors and eye-movement data to conclude that different tasks and levels of experience affect the behavior of students searching for information on the web. In general, user studies and logs are used separately because they are adopted with different aims in mind (Agostii et al., 2009). For instance, Robert Capra, Bill Kules, Matt Banta, and Tito Sierra (2009) describe the use of log data from the online public access catalog (OPAC) to develop a set of grounded tasks. At the same time, through the use of a remote eye tracker in a controlled laboratory setting, they collect eye-tracking data to examine users' behaviors in developing exploratory search tasks. The authors report that data collection using the eye tracker was a difficult process as was using the log data to develop the search tasks.

Gi Woong Yun (2009) differentiates two types of methods for collecting log files: servers and clients. The server-side method is a low-cost, non-intrusive way of collecting data from a large number of individuals with minimal staff involvement. This method uses web log files to identify users' accesses to files in a certain web server. The client-side method requires some contact with study participants because of the need to install a monitoring program on the users' computers. Client-side methods are very invasive, require high staff involvement, and have high costs due to users' recruitments. Gheorghe Muresan (2009) states that data captured by server-side and client-side logging are complementary and typically used to answer different research questions.

To enhance the results of log analysis and test findings, other data-gathering methods can be applied, such as questionnaires, surveys, interviews, or observation studies (Agostii et al., 2009; Black, 2009; Jamali et al., 2005; Kostkova & Madle, 2009). Combining quantitative data -for example, log analysis with qualitative data- allows researchers to cross-check the analysis and fills in knowledge gaps. In addition, this combination provides a much more in-depth picture of how a digital library may be impacting its users' community and their work, and it also explains the information-seeking behavior of the users discovered in the logs. One specific example, presented by Maristella Agostii, Franco Crivellari, and Nick Di Nunzio (2009), concludes that when implicit methods such as users' interaction logs and explicit methods such as users' questionnaires are combined, the results are more scientifically informative than those obtained when the two types of studies are conducted alone. Thus, by incorporating log analysis into our holistic matrix, library managers gain important input on users' behaviors and the possibility of identifying potential failures in the library system at the time of delivering services to their users.

2.3.5 The proposed holistic evaluation matrix

By combining the methodologies discussed in the preceding sections and the conceptual matrix defined by Scott Nicholson, this article proposes a holistic view of the processes, resources, and activities present in libraries from an economic perspective (Figure 2.2). We strongly support the idea that information must be collected from many separate sources, such as library information systems, library statistics, observations, surveys, and users' inquiries, in order to have enough input and different points of view for an adequate decision-making process.

	<i>Library System</i>	<i>Library Collection</i>
<i>Internal Perspective (Library)</i>	1) Service Analysis Processes cost, Time, Resources	4) Usage analysis Transaction log analysis, deep log analysis
<i>External Perspective (Users)</i>	2) Quality Analysis Statistics gathering, Suggestion boxes, Usability testing, Satisfaction surveys	3) Collection Analysis Citations patterns, Publishing patterns, Journals downloaded, and Journals' impact factor

Figure 2.2: Methodologies proposed to economical evaluate a library through a holistic perspective

The approach for implementing this matrix starts by identifying the services or activities involved in libraries and by calculating the costs of different resources (staff, equipment, facilities, collection, etc.). In order to do so, qualitative mechanisms for assessing library effectiveness should be included -for example, observation, interviews, surveys, expert opinions, process analysis, organizational structure analysis, standards, and peer comparison. Quantitative techniques are also required to evaluate efficiency, usefulness, and manipulation of the system. Citation analysis, log analysis, statistics gathering, and stopwatch techniques are useful methods that can be included.

To collect these data, typical data sources could include the following: 1) integrated library systems, which contain information about process performance in the library, circulation data, acquisition, and so on; 2) the library portal used as a front end for the different types of electronic resources; 3) the OPAC as a system to support digital reference services; 4) the interlibrary loan system from consortiums (Nicholson, 2006b); 5) the LibQUAL+® survey system; and 6) information systems for demographic information.

However, some considerations must be taken into account when collecting data from these heterogeneous data sources (Poll, 2001). These factors include 1) lack of well-defined standards for some specific analysis, such as the abbreviation of journal names, access to electronic collection,

and e-lending; 2) the need for a common understanding of what sources and data must be considered; 3) the need for integrating multiple data sources from the library, university, consortiums, and suppliers; 4) differences of requirements between traditional and digital collections (for example, digital libraries require licenses for a certain time period, links to remote resources, or prepaid pay-per-view); and 5) the large volume of data generated by all different sources, for instance, web logs. To develop a structure for a holistic analysis, data generated by multiple data sources must be integrated. Unfortunately, such integration presents a big challenge because these different sources normally use dissimilar formats and access methods (Ying Wah et al., 2007). To overcome these shortcomings, Scott Nicholson (2003b) proposes the aid of a data warehouse to integrate, filter, and process all the information extracted from many different systems based on the holistic matrix.

2.4 Data warehouse architecture for library holistic evaluation

A data warehouse is defined as “a repository of integrated information from distributed, autonomous, and possibly heterogeneous, sources” (quoted in Bleyberg et al. (1999, p. 546)). Based on the measures proposed in this study and the typical structure of a data warehouse (W. H. Inmon, 2005), the resulting system architecture of a library’s data warehouse, as shown in Figure 2.3, is composed of three layers: 1) data source; 2) data extraction, cleansing and storage; and 3) data presentation area.

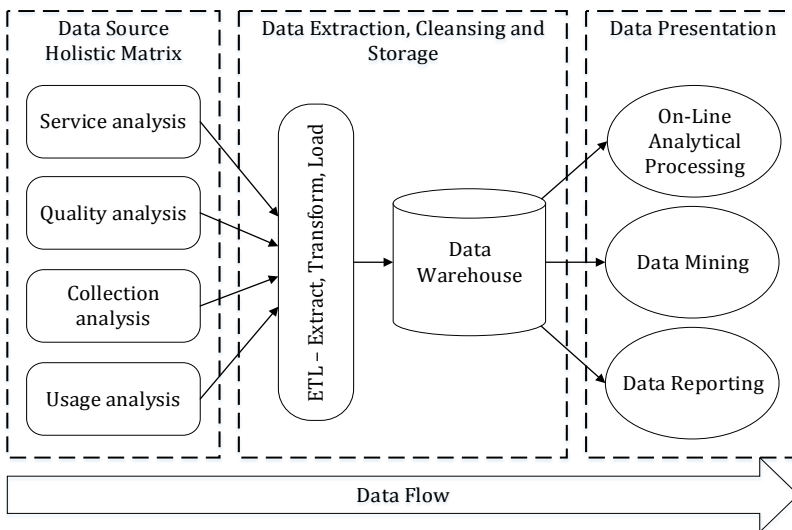


Figure 2.3: Data warehouse architecture for library holistic evaluation

2.4.1 Data source layer

The data source layer is composed of the information extracted from different data sources. In our structure, data sources selected are based on the holistic matrix, which includes the analysis of processes, resources and costs of library services; the point of view of the users on the quality of services; the usefulness of the library collection; and the users’ behaviors in the library system.

2.4.2 Data extraction, cleansing and storage layer

The resulting data are processed by the data extraction, cleansing, and storage layer through extract, transform, load (ETL) processes, allowing a clean, homogeneous, and anonymous version of the library data. ETL is a group of processes whereby the information collected from operative systems is converted into a uniform format required by the data warehouse (Laitinen & Saarti, 2012). ETL also includes tools for loading the data into the data warehouse and for periodically refreshing it. This is a challenging and time-consuming task because the process must combine all the different data sources and convert them into a uniform format, excluding possible inconsistencies, redundancies, and incompatibilities (Nicholson, 2003b). At the same time, ETL processes play a key part in protecting patron privacy during data warehousing (Laitinen & Saarti, 2012).

Once the data have been processed, the next step is to build the data warehouse. Because this process is the most tedious and time-consuming part, Scott Nicholson (2003b) suggests starting with a narrowly specific query, working through the entire process, and then iteratively continuing to develop the data warehouse. This is done in order to minimize the initial time required and also to improve the collection and cleansing algorithms as early as possible.

2.4.3 Data presentation layer

The resulting data are processed by the data extraction, cleansing, and storage layer through extract, transform, load (ETL) processes, allowing a clean, homogeneous, and anonymous version of the library data. ETL is a group of processes whereby the information collected from operative systems is converted into a uniform format required by the data warehouse (Laitinen & Saarti, 2012). ETL also includes tools for loading the data into the data warehouse and for periodically refreshing it. This is a challenging and time-consuming task because the process must combine all the different data sources and convert them into a uniform format, excluding possible inconsistencies, redundancies, and incompatibilities (Nicholson, 2003b). At the same time, ETL processes play a key part in protecting patron privacy during data warehousing (Laitinen & Saarti, 2012).

Once the data have been processed, the next step is to build the data warehouse. Because this process is the most tedious and time-consuming part, Scott Nicholson (2003b) suggests starting with a narrowly specific query, working through the entire process, and then iteratively continuing to develop the data warehouse. This is done in order to minimize the initial time required and also to improve the collection and cleansing algorithms as early as possible.

2.5 Conclusions

Libraries are accustomed to constant evaluation; consequently, they have a long history on data collection statistics (Laitinen & Saarti, 2012). Unfortunately, these statistics are only partially used for decision-making processes because of the wide variety of formats and the lack of efficient methods for grouping information. In this article, a complete framework and set of tools to holistically analyze libraries for financial decisions have been proposed. The approach for implementing the structure is to start extracting and collecting the information generated based on the two-dimension holistic matrix. The theoretical matrix is used to analyze the library collection and services from internal and external perspectives. Furthermore, several methods and appropriate measurement tools have been evaluated and proposed for an integrated decision-making process. Library managers can select one or more instruments in every quadrant based on the current availability or decide to include other measurements and detailed issues in the model. An example of organizing and collecting the information based on this holistic approach is presented by Siguenza-Guzman, Holans, et al (2013). The authors document the preliminary experiences of the implementation, concluding that the holistic model is a simple and powerful

structure for grouping library information. Although, the authors support the practical validity of the proposed approach, they also describe important considerations that need to be borne in mind, for example, the time required to implement the complete approach, as well as the need for dedicated systems to automate the different quadrants.

In addition, this study proposes the architecture of a data warehouse to store the collected data. This resource will allow the use of information, not only in traditional measures or for generating reports but also to enhance decision-making. For instance, information on the following four scenarios is accessible: 1) redistributing and prioritizing the allocation of resources assigned to a specific service; 2) gaining knowledge about users coming into the library and also users who are served by digital services; 3) awareness of the gaps and strengths in services and collections; and 4) the building of collections based on a library's holdings, users priorities, and technological tendencies. Ultimately, this article attempts to integrate this structure with an optimization tool to determine optimal resource allocation decision making in specific scenarios such as budget decreases, journal subscriptions and cancellations, and the creation of new services.

References

- ACRL Research Planning and Review Committee. (2010). 2010 top ten trends in academic libraries. *College & Research Libraries News*, 71(6), 286–292.
- Agee, J. (2005). Collection evaluation: A foundation for collection development. *Collection Building*, 24(3), 92–95. <http://doi.org/10.1108/01604950510608267>
- Agostii, M., Crivellari, F., & Di Nunzio, N. (2009). Evaluation of digital library services using complementary logs. In N. J. Belkin, R. Bierig, G. Buscher, L. van Elst, J. Gwizdka, J. Jose, & J. Teevan (Eds.), *Proceedings of the Workshop on Understanding the User - Logging and Interpreting User Interactions in Information Search and Retrieval* (pp. 32–35). Boston, MA, USA. Retrieved from http://comminfo.rutgers.edu/~jacekg/pubs/txt/2009_UIIR-workshop-proceedings.pdf#page=42
- Allen Press, Inc. (2012). 2012 Study of Subscription Prices for Scholarly Society Journals: Society Journal Pricing Trends and Industry Overview (p. 20 p.). Allen Press, Inc. Retrieved from http://allenpress.com/system/files/pdfs/library/2012_AP_JPS.pdf
- Association of Research Libraries. (2012). LibQUAL+®: Charting Library Service Quality. Retrieved from http://www.libqual.org/about/about_lq/general_info
- Beile, P. M., Boote, D. N., & Killingsworth, E. K. (2004). A Microscope or a Mirror?: A Question of Study Validity Regarding the Use of Dissertation Citation Analysis for Evaluating Research Collections. *The Journal of Academic Librarianship*, 30(5), 347–353. <http://doi.org/10.1016/j.jacalib.2004.06.001>
- Bertot, J. C. (2011). Concluding Comments: 2010 Library Assessment Conference. *The Library Quarterly*, 81(1), 127–128. <http://doi.org/10.1086/657502>
- Black, E. L. (2009). Web Analytics: A Picture of the Academic Library Web Site User. *Journal of Web Librarianship*, 3(1), 3–14. <http://doi.org/10.1080/19322900802660292>
- Bland, R. N. (1980). The college textbook as a tool for collection evaluation, analysis, and retrospective collection development. *Library Acquisitions: Practice & Theory*, 4(3–4), 193–197. [http://doi.org/10.1016/0364-6408\(80\)90019-8](http://doi.org/10.1016/0364-6408(80)90019-8)
- Bleyberg, M. Z., Zhu, D., Cole, K., Bates, D., & Zhan, W. (1999). Developing an integrated library decision support data warehouse. In 1999 IEEE International Conference on Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings (Vol. 2, pp. 546–551). Tokyo, Japan: IEEE. <http://doi.org/10.1109/ICSMC.1999.825319>
- Blixrud, J. C. (2003). Assessing library performance: new measures, methods, and models. In *Proceedings of the IATUL Conferences* (p. Paper 9). Ankara, Turkey: Purdue e-Pubs. Retrieved from <http://docs.lib.purdue.edu/iatul/2003/papers/9/>
- Brook, J., & Salter, A. (2012). E-books and the Use of E-book Readers in Academic Libraries: Results of an Online Survey. *Georgia Library Quarterly*, 49(4). Retrieved from <http://digitalcommons.kennesaw.edu/glq/vol49/iss4/10>
- Buchanan, R. A. (2006). Accuracy of Cited References: The Role of Citation Databases. *College & Research Libraries*, 67(4), 292–303.

- Capra, R., Kules, B., Banta, M., & Sierra, T. (2009). Faceted Search for Library Catalogs: Developing Grounded Tasks and Analyzing Eye-Tracking Data. In N. J. Belkin, R. Bierig, G. Buscher, L. van Elst, J. Gwizdka, J. Jose, & J. Teevan (Eds.), *Proceedings of the Workshop on Understanding the User - Logging and Interpreting User Interactions in Information Search and Retrieval* (pp. 16–18). Boston, MA, USA. Retrieved from http://comminfo.rutgers.edu/~jacekg/pubs/txt/2009_UIIR-wrkshp-proceedings.pdf#page=26
- Chan, G. R. Y. C. (2008). Aligning collections budget with program priorities: A modified zero-based approach. *Library Collections, Acquisitions, and Technical Services*, 32(1), 46–52. <http://doi.org/10.1016/j.lcats.2008.06.001>
- Ching, S. H., Leung, M. W., Fidow, M., & Huang, K. L. (2008). Allocating costs in the business operation of library consortium: The case study of Super e-Book Consortium. *Library Collections, Acquisitions, and Technical Services*, 32(2), 97–103. <http://doi.org/10.1016/j.lcats.2008.08.013>
- Cook, C. (2002). The maturation of assessment in academic libraries: the role of LibQUAL+TM. *Performance Measurement and Metrics*, 3(2). Retrieved from <http://www.emeraldinsight.com/journals.htm?issn=1467-8047&volume=3&issue=2&articleid=1491221&show=html>
- Cooper, R., & Kaplan, R. S. (1988). Measure costs right: Make the right decision. *Harvard Business Review*, 66(5), 96–103.
- Cottrell, T. (Terry). (2012). Three phantom budget cuts and how to avoid them. *Bottom Line: Managing Library Finances*, The, 25(1), 16–20. <http://doi.org/10.1108/08880451211229171>
- Cox, C. (2010). Proposed Consolidation of Branch Libraries. Retrieved from <http://library.wvu.edu/dean/proposed-consolidation-branch-libraries>
- Creaser, C. (2006). One size does not fit all: User surveys in academic libraries. *Performance Measurement and Metrics*, 7(3), 153–162. <http://doi.org/10.1108/14678040610713110>
- Davis, P. M., & Solla, L. R. (2003). An IP-level analysis of usage statistics for electronic journals in chemistry: Making inferences about user behavior. *Journal of the American Society for Information Science and Technology*, 54(11), 1062–1068. <http://doi.org/10.1002/asi.10302>
- Diodato, V. P. (1994). *Dictionary of Bibliometrics* (1 edition). New York, NY, USA: The Haworth Press, Inc.
- Duy, J., & Vaughan, L. (2006). Can electronic journal usage data replace citation data as a measure of journal use? An empirical examination. *The Journal of Academic Librarianship*, 32(5), 512–517. <http://doi.org/10.1016/j.acalib.2006.05.005>
- Editors of the American Heritage Dictionaries. (2011). The American Heritage Dictionary entry: holistic [Dictionary]. Retrieved from <http://www.ahdictionary.com/word/search.html?q=holistic&submit.x=56&submit.y=19>
- Edwards, S. (1999). Citation Analysis as a Collection Development Tool: A Bibliometric Study of Polymer Science Theses and Dissertations. *Serials Review*, 25(1), 11–20.
- Ellis-Newman, J. (2003). Activity-Based Costing in user services of an academic library. *Library Trends*, 51(3), 333–348.
- Ellis-Newman, J., Izan, H., & Robinson, P. (1996). Costing support services in universities: An application of activity-based costing. *Journal of Institutional Research in Australasia*, 5(1), 75–86.
- Ellis-Newman, J., & Robinson, P. (1998). The cost of library services: Activity-based costing in an Australian academic library. *The Journal of Academic Librarianship*, 24(5), 373–379. [http://doi.org/10.1016/S0099-1333\(98\)90074-X](http://doi.org/10.1016/S0099-1333(98)90074-X)
- Enger, K. B. (2009). Using citation analysis to develop core book collections in academic libraries. *Library & Information Science Research*, 31(2), 107–112. <http://doi.org/10.1016/j.lisr.2008.12.003>
- Ernst, D. J., & Segall, P. (1995). Information Resources and Institutional Effectiveness: The Need for a Holistic Approach to Planning and Budgeting. *Cause/Effect*, 18(1), 11–13.
- Feyereisen, P., & Spoiden, A. (2009). Can Local Citation Analysis of Master's and Doctoral Theses help Decision-Making about the Management of the Collection of Periodicals? A Case

- Study in Psychology and Education Sciences. *The Journal of Academic Librarianship*, 35(6), 514–522. <http://doi.org/10.1016/j.acalib.2009.08.018>
- Franklin, B., Kyrillidou, M., & Plum, T. (2009). From usage to user: library metrics and expectations for the evaluation of digital libraries. In G. Tsakonas & C. Papatheodorou (Eds.), *Evaluation of Digital Libraries: an insight into useful applications and methods* (pp. 17–39). Oxford, UK: Chandos. Retrieved from <http://www.libqual.org/documents/admin/Ch2-EvalofDigitalLib.pdf>
- Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., Kapidakis, S., ... Sølvyberg, I. (2007). Evaluation of digital libraries. *International Journal on Digital Libraries*, 8(1), 21–38. <http://doi.org/10.1007/s00799-007-0011-z>
- Gerdson, T. (2002). Activity Based Costing as a Performance Tool for Library & Information Technology Services. In *Proceedings of the 4th Northumbria International Conference on Performance Measurement in Libraries Information Services* (Vol. 4, pp. 119–126). Washington, DC, USA: Association of Research Libraries.
- Goddard, A., & Ooi, K. (1998). Activity-Based Costing and central overhead cost allocation in universities: A case study. *Public Money and Management*, 18(3), 31–38. <http://doi.org/10.1111/1467-9302.00124>
- Gonçalves, M., Luo, M., Shen, R., Ali, M., & Fox, E. (2002). An XML Log Standard and Tool for Digital Library Logging Analysis. In M. Agosti & C. Thanos (Eds.), *Research and Advanced Technology for Digital Libraries* (Vol. 2458, pp. 215–216). Springer Berlin / Heidelberg. Retrieved from <http://www.springerlink.com/content/6d5505p4whkn4mvl/abstract/>
- Guarria, C. I. (2009). How using an allocation formula changed funding allocations at Long Island University. *Collection Building*, 28(2), 44–50. <http://doi.org/10.1108/01604950910953071>
- Guarria, C. I., & Wang, Z. (2011). The economic crisis and its effect on libraries. *New Library World*, 112(5/6), 199–214. <http://doi.org/10.1108/03074801111136248>
- Heaney, M. (2004). Easy as ABC? Activity-based costing in Oxford University Library Services. *Bottom Line: Managing Library Finances*, The, 17(3), 93–97. <http://doi.org/10.1108/08880450410699701>
- Hernon, P., & Altman, E. (1996). *Service quality in academic libraries*. Norwood, NJ, USA: Ablex Publishing Corporation.
- Hernon, P., & Altman, E. (2010). *Assessing Service Quality: Satisfying the Expectations of Library Customers*, 2nd Ed. Chicago, IL, USA: American Library Association. Retrieved from http://books.google.be/books?hl=en&lr=&id=TsGGXWRh1yoC&oi=fnd&pg=PR5&dq=Hernon+quality&ots=DWERiYeUH6&sig=TujMv3FOASvC_V8p0-JPv2Lb07M
- He, Y., & Cheung Hui, S. (2002). Mining a Web Citation Database for author co-citation analysis. *Information Processing & Management*, 38(4), 491–508. [http://doi.org/10.1016/S0306-4573\(01\)00046-2](http://doi.org/10.1016/S0306-4573(01)00046-2)
- Horn, A., & Owen, S. (2009). Mind the gap 2014 : research to inform the next five years of library development. In *Innovate, collaborate : conference proceedings EDUCAUSE Australasia 2009* (pp. 1–12). Perth, Western Australia. Retrieved from <http://dro.deakin.edu.au/view/DU:30016487>
- Hudomalj, E., & Vidmar, G. (2003). OLAP and bibliographic databases. *Scientometrics*, 58(3), 609–622. <http://doi.org/10.1023/B:SCIE.0000006883.28709.d2>
- Hwang, S.-Y., Keezer, P., & O'Neill, E. T. (2003). The bibliomining process: Data warehousing and data mining for libraries. Sponsored by SIG LT. *Proceedings of the American Society for Information Science and Technology*, 40(1), 478–479. <http://doi.org/10.1002/meet.1450400184>
- Inmon, W. H. (2005). *Building the Data Warehouse* (4th ed.). Indianapolis, IN, USA: Wiley.
- Jamali, H. R., Nicholas, D., & Huntington, P. (2005). The use and users of scholarly e-journals: a review of log analysis studies. *Aslib Proceedings*, 57(6), 554–571. <http://doi.org/10.1108/00012530510634271>
- Kao, S.-C., Chang, H.-C., & Lin, C.-H. (2003). Decision support for the academic library acquisition budget allocation via circulation database mining. *Information Processing & Management*, 39(1), 133–147. [http://doi.org/10.1016/S0306-4573\(02\)00019-5](http://doi.org/10.1016/S0306-4573(02)00019-5)
- Kaplan, R. S., & Anderson, S. R. (2003). Time-Driven Activity-Based Costing. SSRN eLibrary. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=485443

- Kaplan, R. S., & Anderson, S. R. (2007). *Time-Driven Activity-Based Costing: A simpler and more powerful path to higher profits*. Boston, MA, USA: Harvard Business School Press.
- Kostkova, P., & Madle, G. (2009). User-Centered Evaluation Model for Medical Digital Libraries. In D. Riaño (Ed.), *Knowledge Management for Health Care Procedures* (Vol. 5626, pp. 92–103). Springer Berlin / Heidelberg. Retrieved from <http://www.springerlink.com/content/4736023260127808/abstract/>
- Kowalski, G. (2011). Chapter 9: Information System Evaluation. In *Information Retrieval Architecture and Algorithms* (pp. 253–281). Springer US. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4419-7716-8_9
- Kruschwitz, U., Albakour, M.-D., Niu, J., Leveling, J., Nanas, N., Kim, Y., ... De Roeck, A. (2011). Moving towards Adaptive Search in Digital Libraries. In R. Bernardi, S. Chambers, B. Gottfried, F. Segond, & I. Zaihrayeu (Eds.), *Advanced Language Technologies for Digital Libraries* (Vol. 6699, pp. 41–60). Springer Berlin / Heidelberg. Retrieved from <http://www.springerlink.com/content/f8gj7410qwxw3065/abstract/>
- Lafouge, T., & Lainé-Cruzel, S. (1997). A new explanation of the geometric law in the case of library circulation data. *Information Processing & Management*, 33(4), 523–527. [http://doi.org/10.1016/S0306-4573\(96\)00075-1](http://doi.org/10.1016/S0306-4573(96)00075-1)
- Laitinen, M., & Saarti, J. (2012). A model for a library-management toolbox: Data warehousing as a tool for filtering and analyzing statistical information from multiple sources. *Library Management*, 33(4/5), 253–260. <http://doi.org/10.1108/01435121211242290>
- Lancaster, F. W. (1977). *The Measurement and Evaluation of Library Services*. Washington, D.C.: Information Resources Press.
- Lancaster, F. W. (1988). *If You Want to Evaluate Your Library...* Champaign, IL: Champaign: University of Illinois Press.
- Linn, M. (2007). Budget systems used in allocating resources to libraries. *Bottom Line: Managing Library Finances*, The, 20(1), 20–29. <http://doi.org/10.1108/08880450710747425>
- Matthews, J. R. (2011). Assessing Organizational Effectiveness: The Role of Performance Measures. *The Library Quarterly*, 81(1), 83–110. <http://doi.org/10.1086/657447>
- Matthews, J. R. (2013). Valuing Information, Information Services, and the Library: Possibilities and Realities. *Portal: Libraries and the Academy*, 13(1), 91–112.
- McKendrick, J. (2011). *Funding and Priorities: The Library Resource Guide Benchmark Study on 2011 Library Spending Plans* (p. 40). Chatham, NJ, USA: Unisphere Research, a division of Information Today, Inc. Retrieved from <http://conan.lib.muohio.edu/ebooks/Fundingandpriorities.pdf>
- Mehrubeoglu, M., Pham, L. M., Le, H. T., Muddu, R., & Ryu, D. (2011). Real-time eye tracking using a smart camera. In 2011 IEEE Applied Imagery Pattern Recognition Workshop (AIPR) (pp. 1–7). <http://doi.org/10.1109/AIPR.2011.6176373>
- Muresan, G. (2009). An Integrated Approach to Interaction Design and Log Analysis. In B. J. Jansen, A. Spink, & I. Taksa, *Handbook of Research on Web Log Analysis* (pp. 227–255). New York, NY, USA: Information Science Reference (an imprint of IGI Global).
- Nicholas, D., Huntington, P., & Homewood, J. (2003). Assessing used content across five digital health information services using transaction log files. *Journal of Information Science*, 29(6), 499–515. <http://doi.org/10.1177/0165551503296007>
- Nicholas, D., Huntington, P., Jamali, H. R., & Tenopir, C. (2006). What deep log analysis tells us about the impact of big deals: case study OhioLINK. *Journal of Documentation*, 62(4), 482–508. <http://doi.org/10.1108/00220410610673864>
- Nicholas, D., Huntington, P., & Watkinson, A. (2003). Digital journals, Big Deals and online searching behaviour: a pilot study. *Aslib Proceedings*, 55(1/2), 84–109. <http://doi.org/10.1108/00012530310462742>
- Nicholas, D., Huntington, P., & Watkinson, A. (2005). Scholarly journal usage: the results of deep log analysis. *Journal of Documentation*, 61(2), 248–280. <http://doi.org/10.1108/00220410510585214>
- Nicholas, D., Huntington, P., & Williams, P. (2002). Evaluating metrics for comparing the use of web sites: a case study of two consumer health web sites. *Journal of Information Science*, 28(1), 63–75. <http://doi.org/10.1177/016555150202800107>
- Nicholas, D., Rowlands, I., Jubb, M., & Jamali, H. R. (2010). The impact of the economic downturn on libraries: With special reference to university libraries. *The Journal of Academic Librarianship*, 36(5), 376–382. <http://doi.org/10.1016/j.acalib.2010.06.001>

- Nicholson, S. (2003). The bibliomining process: Data Warehousing and Data Mining for library decision-making. *Information Technology and Libraries*, 22(4), 146–151.
- Nicholson, S. (2004). A conceptual framework for the holistic measurement and cumulative evaluation of library services. *Journal of Documentation*, 60(2), 164–182. <http://doi.org/10.1108/00220410410522043>
- Nicholson, S. (2006a). Approaching librarianship from the data: Using bibliomining for evidence-based librarianship. *Library Hi Tech*, 24(3), 369–375. <http://doi.org/10.1108/07378830610692136>
- Nicholson, S. (2006b). The basis for bibliomining: Frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services. *Information Processing & Management*, 42(3), 785–804. <http://doi.org/10.1016/j.ipm.2005.05.008>
- Nicholson, S., & Stanton, J. M. (2003). Gaining strategic advantage through Bibliomining: Data Mining for management decisions in corporate, special, digital, and traditional libraries. In H. R. Nemati & C. D. Barko, *Organizational data mining: Leveraging enterprise data resources for optimal performance* (pp. 247–262). Hershey, PA, USA: Idea Group Publishing (an imprint of Idea Group Inc.). Retrieved from http://arizona.openrepository.com/arizona/bitstream/10150/106383/1/Nicholson_3.pdf
- Nitecki, D. A., & Hernon, P. (2000). Measuring service quality at Yale university's libraries. *The Journal of Academic Librarianship*, 26(4), 259–273. [http://doi.org/10.1016/S0099-1333\(00\)00117-8](http://doi.org/10.1016/S0099-1333(00)00117-8)
- Novak, D. D., Paulos, A., & Clair, G. S. (2011). Data-driven budget reductions: A case study. *The Bottom Line: Managing Library Finances*, 24(1), 24–34. <http://doi.org/10.1108/08880451111142015>
- Pernot, E., Roodhooft, F., & Van den Abbeele, A. (2007). Time-Driven Activity-Based Costing for inter-library services: A case study in a university. *The Journal of Academic Librarianship*, 33(5), 551–560. <http://doi.org/10.1016/j.jacalib.2007.06.001>
- Peters, T. A. (1993). The history and development of transaction log analysis. *Library Hi Tech*, 11(2), 41–66. <http://doi.org/10.1108/eb047884>
- Peters, T. A. (2001). What's the Big Deal? *Journal of Academic Librarianship*, 27(4), 302.
- Poll, R. (2001). Performance measures for library networked services and resources. *Electronic Library*, 19(5), 307–315. <http://doi.org/10.1108/02640470110408887>
- Saito, H., Terai, H., Egusa, Y., Takaku, M., Miwa, M., & Kando, N. (2009). How Task Types and User Experiences Affect Information-Seeking Behavior on the Web: Using Eye-tracking and Client-side Search Logs. In N. J. Belkin, R. Bierig, G. Buscher, L. van Elst, J. Gwizdzka, J. Jose, & J. Teevan (Eds.), *Proceedings of the Workshop on Understanding the User - Logging and Interpreting User Interactions in Information Search and Retrieval* (pp. 19–22). Boston, MA, USA. Retrieved from http://comminfo.rutgers.edu/~jacekg/pubs/txt/2009_UIIR-wrkshp-proceedings.pdf#page=11
- Saunders, E. S. (2007). The LibQUAL Phenomenon: Who Judges Quality? *Libraries Research Publications*, 47, 21–24.
- Siguenza-Guzman, L., Holans, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2013). Towards a holistic analysis tool to support decision-making in libraries. In *Proceedings of the IATUL Conferences*. Cape Town, South Africa: Purdue e-Pubs. Retrieved from <http://docs.lib.purdue.edu/iatul/2013/papers/29/>
- Siguenza-Guzman, L., Saquicela, V., & Cattrysse, D. (2014). Design of an Integrated Decision Support System for library holistic evaluation. In *Proceedings of IATUL Conferences* (pp. 1–12). Espoo, Finland.
- Siguenza-Guzman, L., Van Den Abbeele, A., & Cattrysse, D. (2014). Time-Driven Activity-Based Costing Systems for Cataloguing Processes: A Case Study. *LIBER Quarterly*, 23(3), 160–186.
- Siguenza-Guzman, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2013). Recent evolutions in costing systems: A literature review of Time-Driven Activity-Based Costing. *ReBEL - Review of Business and Economic Literature*, 58(1), 34–64.
- Siguenza-Guzman, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2014). Using Time-Driven Activity-Based Costing to Support Library Management Decisions: A

- Case Study for Lending and Returning Processes. *Library Quarterly: Information, Community, Policy*, 84(1), 1–23.
- Skilbeck, M., & Connell, H. (2001). Activity based costing: a study to develop a costing methodology for Library and Information Technology activities for the Australian higher education sector (p. 124 p.). Information and Education Services Division, University of Newcastle. Retrieved from <http://www.dest.gov.au/NR/rdonlyres/49D61919-B87F-4369-9AA3-4DE5876FB32B/4653/libraries.pdf>
- Smith, D. A. (2008). Percentage based allocation of an academic library materials budget. *Collection Building*, 27(1), 30–34. <http://doi.org/10.1108/01604950810846224>
- Spreckelsen, C., Deserno, T. M., & Spitzer, K. (2011). Visibility of medical informatics regarding bibliometric indices and databases. *BMC Medical Informatics and Decision Making*, 11(1), 24. <http://doi.org/10.1186/1472-6947-11-24>
- Stouthuysen, K., Swiggers, M., Reheul, A.-M., & Roodhooft, F. (2010). Time-Driven Activity-Based Costing for a library acquisition process: A case study in a Belgian University. *Library Collections, Acquisitions, and Technical Services*, 34(2-3), 83–91. <http://doi.org/16/j.lcats.2010.05.003>
- Sudarsan, P. K. (2006). A resource allocation model for university libraries in India. *Bottom Line: Managing Library Finances, The*, 19(3), 103–110. <http://doi.org/10.1108/08880450610682527>
- Wallace, D. P., & Van Fleet, C. J. (2001). *Library Evaluation: A Casebook and Can-Do Guide*. Englewood, Colorado, EEUU: Libraries Unlimited.
- Walters, W. H. (2013). E-books in Academic Libraries: Challenges for Acquisition and Collection Management. *Portal : Libraries and the Academy*, 13(2), 187.
- Woong Yun, G. (2009). The Unit of Analysis and the Validity of Web Log Data. In B. J. Jansen, A. Spink, & I. Taksa, *Handbook of Research on Web Log Analysis* (pp. 165–180). New York, NY, USA: Information Science Reference (an imprint of IGI Global).
- Wright, S., & White, L. S. (2007). *Library Assessment: SPEC Kit 303*. Association of Research Libraries, 14.
- Yilmaz, R. (2008). Creating the profit focused organization using Time-Driven Activity Based Costing. In EABR & TLC Conferences Proceedings (p. 8). Salzburg, Austria: Clute Institute for Academic Research. Retrieved from <http://rifatyilmaz.net/SLZ08-191.pdf>
- Ying Wah, T., Hooi Peng, N., & Sue Hok, C. (2007). Building Data Warehouse. In *Proceedings of the 24th South East Asia Regional Computer Conference*. Bangkok, Thailand. Retrieved from http://www.ijcim.au.edu/SpecialEditions/v15nSP4/P05SEARCC_BuildingDataWarehous e.pdf
- Zeithaml, V. A., Parasuraman, A., & Berry, L. L. (1990). *Delivering Quality Service: Balancing Customer Perceptions and Expectations*. Simon and Schuster.
- Zhang, Y. (2010). Developing a holistic model for digital library evaluation. *Journal of the American Society for Information Science and Technology*, 61(1), 88–110. <http://doi.org/10.1002/asi.21220>
- Zipp, L. S. (1996). Thesis and Dissertation Citations as Indicators of Faculty Research Use of University Library Journal Collections. *Library Resources & Technical Services*, 40(4), 335–342.

Chapter 3: **Literature Review of Time-Driven Activity-Based Costing**

Siguenza Guzman, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., Cattrysse, D. (2013). Recent evolutions in costing systems: A literature review of Time-Driven Activity-Based Costing. Review of Business and Economic Literature, 58 (1), 34-64.

This chapter complements the holistic structure by analyzing a costing system, fundamental to calculate the costs of processes and library services. Specifically, a comprehensive literature review of Time-Driven Activity Based Costing (TDABC) is provided, as well as potential benefits and difficulties encountered during the case study implementation.

Apart from typographical adjustments, the content of this chapter is identical to the content of the published paper quoted above; where necessary, additional information or remarks are added in footnotes. The layout is adapted for consistency throughout this dissertation. Some redundancy with other chapters is unavoidable as an academic article needs its own introductory sections. This, however, entails the advantage that the chapter can be read separately.

Abstract

This article provides a comprehensive literature review of Time-Driven Activity Based Costing (TDABC), a relatively new tool to improve the cost allocation to products and services. After a brief overview of traditional costing and activity based costing systems (ABC), a detailed description of the TDABC model is given and a comparison made between this methodology and its predecessor ABC. Thirty-six empirical contributions using TDABC over the period 2004-2012 were reviewed. The results and conclusions of these studies are grouped according to the main areas of application of the method such as logistics, manufacturing, services, health, hospitality and nonprofit services. Potential benefits, and challenges are identified.

Contributions of the first author

The first author's contributions are: the comprehensive literature study on costing systems, with special emphasis on TDABC, analysis and classification of the case studies sorted by area, description of benefits, challenges and opportunities, as well as the summary of conclusions.

3.1 Introduction

Calculating the cost of products or services remains a difficult exercise, especially in highly competitive environments where in order to guarantee long-term profitability, companies must ensure that their product and service costs should not exceed market prices (Hoozée, Vermeire, & Bruggeman, 2009). However, also in the non-profit and public sector accurate cost estimations are crucial given the need to constantly prioritize spending and to minimize costs because of the limited resources and budget pressures (Linn, 2007; Sudarsan, 2006; Wise & Perushek, 1996).

Costing systems help companies determine the cost of a cost object such as a product or service. Direct costs such as direct labor and materials are relatively easy to measure and can be directly attributed to specific products or services. On the contrary, indirect costs such as marketing, depreciation, training and electricity are not directly attributable to a cost object. Indirect costs are therefore allocated to a cost object using an allocation approach.

TDABC is a cost allocation approach developed by Kaplan and Anderson in 2004 to better attribute the indirect costs to the cost objects in order to obtain more accurate information to set priorities for process improvements, product variety, price settings, and customer relationships. This paper provides an overview of the recent literature on TDABC. First, the predecessors of TDABC such as traditional costing systems and ABC systems are briefly discussed. A detailed description of the TDABC model is then presented, followed by an extensive comparison between TDABC and its main predecessor ABC.

The first aim of this paper is to analyze thirty-six case studies carried out with TDABC over the period 2004-2012. This information is categorized in a way that provides a useful understanding of how TDABC has been applied in specific areas. The second aim is to use the literature review as a platform to identify potential benefits and difficulties encountered by researchers when applying TDABC in real cases. From these results some conclusions and suggestions can be drawn.

3.2 Costing Systems

3.2.1 Traditional costing systems

Robert Kaplan and Robin Cooper (1998) analyzed several integrated cost systems to drive profitability and performance. One is the *traditional costing system* used mainly in the past and now merely for financial reporting procedures. In a traditional costing system, direct costs such as direct labor and materials, are directly attributed to the cost objects. On the contrary, indirect costs such as marketing, depreciation, training and electricity are typically allocated to each cost object using a single or a few volume-based cost drivers (e.g. direct labor, machine hours or units of output). This type of costing system was created when companies manufactured products with little variety and a predominant proportion of direct costs; or when supporting activities and its accompanying indirect costs were limited (Novičević & Ljilja, 1999). Presently, traditional costing systems still work well in stable environments with small or fixed indirect costs and little variation in activities, products or services (Kaplan & Cooper, 1998; Tse & Gong, 2009).

However, because of automation, short product life cycles and high products and services variety, most production and service environment have changed. Therefore, the cost system that was adequate for homogeneous cost pools driven by a single cost rate could now be given distorted signals about profitability and performance when using volume-based allocation rates. The limitation of traditional costing systems is that they are unable to allocate the indirect costs of many resources of a company (i.e. specific costs related to marketing, research, depreciation, support, training, electricity) in an accurate way (Kaplan & Cooper, 1998; Yilmaz, 2008a). Since indirect costs have become increasingly more important than direct costs and those costs are not accurately attributed to the different activities and products, traditional costing systems are unable to

estimate adequate cost information for most organizations today (Ellis-Newman & Robinson, 1998).

3.2.2 Activity-Based Costing System

With the rise of the complexity of companies' operations, the weakness of traditional volume-based costing models becomes more evident (Tse & Gong, 2009). Managers have sought other ways of obtaining more accurate information about costs, being ABC one of the most prominent alternatives (Kaplan & Cooper, 1998). ABC was first developed by practitioners and then introduced in some Harvard Business School teaching cases (Bjørnenak & Mitchell, 2002). It was especially promoted by Cooper and Kaplan in the mid-1980s⁴ (Cooper & Kaplan, 1988; Kaplan & Cooper, 1998).

Because ABC was first designed for manufacturing processes (Gunasekaran & Sarhadi, 1998; Wegmann, 2010), the theory of its promoters is based on the assumption that products differ in the complexity of manufacturing and that the consumption of activities is also in different proportions. Compared to traditional costing methods, ABC is a process which provides a more accurate and efficient management of activity costs since it draws indirect costs more closely to the different activities (Ellis-Newman & Robinson, 1998). Figure 3.1 presents the stages that ABC uses for cost allocation, the cost drivers and the relationships between the resources, activities and objects.

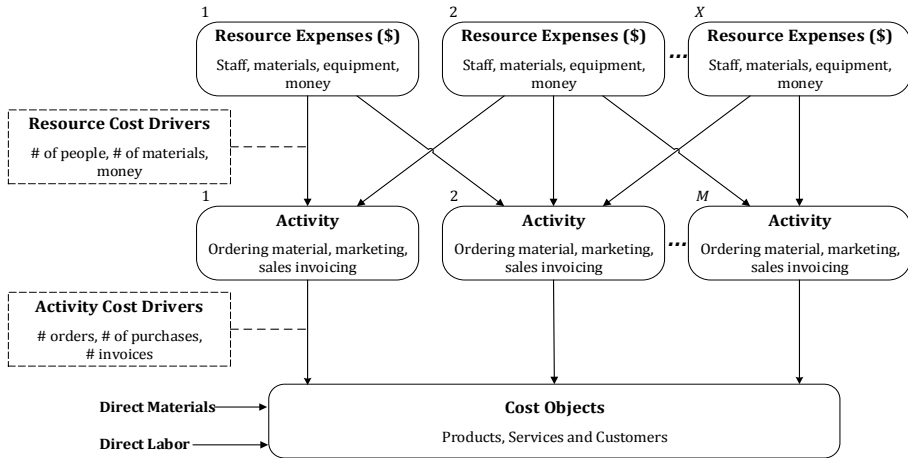


Figure 3.1: Activity-Based Costing Structure (Kaplan & Cooper, 1998)

According to Kaplan and Cooper (1998), ABC is composed of: 1) *Activities* such as ordering materials, marketing, sales invoicing, that consume *resources* (e.g. people, materials, equipment, money). Resource expenses are linked to the different activities through the use of resource cost drivers. A resource cost driver indicates the amount of resources an activity requires. 2) *Objects* such as products or services that require different activities. Activity costs are linked to cost objects using activity cost drivers. An activity cost driver indicates the number of activities an object utilizes. Therefore, resource cost drivers and activity cost drivers are used as a linkage among resources, activities and cost objects.

The information required for an ABC system is collected through interviews, surveys, observations, etc. of the time percentage that employees spend on their several activities. According to Michalska

⁴ For more information about ABC development see Jones and Dugdale (2002) and Gosselin (2006)

and Szewieczek (2007), it is recommended to divide activity costs as detailed as possible, starting first from the more general activities and then going deeply to more detailed aspects.

Since its creation ABC has been used to achieve better accuracy and to increase profits in manufacturing as well as in services, public sector and non-profit setting (Varila, Seppänen, & Suomala, 2007). The use of ABC enhances the traditional costing method contribution, giving the possibility of: 1) including more detailed cost of activities (e.g. direct and indirect costs pools) (Kaplan & Cooper, 1998); 2) getting benefit from a number of indirect-cost pools (e.g. indirect labor cost, indirect material cost, insurance cost, administration cost) (Özbayrak, Akgün, & Türker, 2004); 3) lowering costs through the identification of high cost activities (Acorn Systems, 2007); 4) permitting the detection of unprofitable products, services, customers and useless costs (Acorn Systems, 2007; Kaplan & Cooper, 1998; Yilmaz, 2008a); 5) allowing to identify inefficient or unnecessary activities (Michalska & Szewieczek, 2007; Yilmaz, 2008a); 6) better understanding the origin of costs (Dalci, Tanis, & Kosan, 2010).

Despite these advantages of ABC over traditional costing systems (Clarke, Hill, & Stevens, 1999; Cooper & Kaplan, 1988, 1991; Innes, Mitchell, & Sinclair, 2000), several authors have recognized that ABC models are not the real solution (Dalci et al., 2010; Demeere, Stouthuysen, & Roodhooft, 2009; Everaert, Bruggeman, & De Creus, 2008; Everaert, Bruggeman, Sarens, Anderson, & Levant, 2008; Kaplan & Anderson, 2004, 2007a; Tse & Gong, 2009; Wegmann & Nozile, 2009). For instance:

- The complexity of the actual services or activities is not captured by ABC because of the degree of subjectivity involved in estimating employees' proportion of time spent on each activity.
- The accuracy of data is biased or distorted, because during the interviews employees tend to ignore their idle or unused time; Demeere et al. (2009) also remark that employees will conveniently supply the information based on how it might be used in the future.
- The time, resources and money for data collection are excessive due to the need to re-interview and resurvey people every time an activity or service is changed, updated or removed.
- The cost driver rate is inaccurate because it is calculated assuming that all committed resources are working to full capacity instead of a practical capacity.
- The computational demand required for storing and processing data is very high because it rises non-linearly if ABC needs to be expanded to reflect more granularity and detail on activities.
- The integration between ABC systems and other organizational information systems is limited.
- The use of a single driver rate for each activity makes it difficult to model multi-driver activities.

3.3 Time-Driven Activity-Based Costing System

TDABC is an approach developed by Kaplan and Anderson in order to overcome the difficulties presented in ABC systems (Kaplan & Anderson, 2004). Although ABC has the capability of using time as a cost driver in this new version of ABC, time plays a different role in allocating activity costs to cost objects (Hoozée et al., 2009). For each activity, *costing equations* are calculated based on the time required to perform a transactional activity (Yilmaz, 2008a).

3.3.1 Brief history

Despite the fact that the term TDABC first appeared in 2004, the idea really originated in 1997 (Kaplan & Anderson, 2007b). On the one hand, Steven R. Anderson and his company Acorn Systems began experimenting more accurately with the use of time equations and average time estimates

(Hudig, 2007). These equations were already fed with information gathered from transaction files of an Enterprise Resource Planning (ERP) system. On the other hand and almost simultaneously, Robert S. Kaplan started thinking about capacity and time as improved concepts for ABC models (Hudig, 2007). For instance, Kaplan proposed the idea that an entire cost system could be built based on two parameters: 1) the cost rate for supplying capacity and 2) the capacity used by each transaction (Kaplan & Cooper, 1998).

In 2001, Kaplan joined Acorn Systems to collaborate with Anderson and improve their approach (Kaplan & Anderson, 2007b). Through several discussions, the idea of integrating Anderson's process time equations with Kaplan's capacity planning vision emerged (Hudig, 2007; Kaplan & Anderson, 2007b). Finally in 2004, Kaplan and Anderson introduced TDABC seeking to remedy ABC pitfalls (Kaplan & Anderson, 2004).

3.3.2 The model

TDABC as its predecessor ABC, starts by estimating the cost of supplying capacity (Demeere et al., 2009). However, TDABC estimates resource usage by means of time equations to determine the time needed to perform each activity (Hoozée & Bruggeman, 2010). TDABC assigns resource costs directly to the cost objects using only two parameters: 1) the cost per time unit of supplying resource capacity and 2) an estimate of the time units required to perform a process, an activity or a service.

The first parameter is gathered by dividing the total cost of supplying resource capacity by the practical capacity. The total cost is defined as the cost of all the resources supplied to this department or process (resources such as personnel, supervision, equipment, technology, and infrastructure). The practical capacity is defined as the amount of time that employees work without idle time (Kaplan & Anderson, 2007a). There are two ways to obtain this value: 1) a percentage of the theoretical capacity: assuming the practical capacity is about 80% for people (because of breaks, arrival and departure, training, and meetings), and 85% for machines (because of maintenance, repair, and scheduling fluctuations) of theoretical full capacity; and 2) calculating the real values adjusted for the company. The second number can be obtained through interviews or by direct observation from employees when performing their work; no additional surveys are required. Authors argue that precision is not critical, that a rough accuracy is sufficient because gross inaccuracies will be revealed either in unexpected surpluses or shortages of committed resources (Kaplan & Anderson, 2007b).

Figure 3.2 presents the stages that TDABC uses for cost allocation. Resource expenses are allocated into the activities through the use of resource cost drivers where the unit cost per resource pool is equal to the total cost divided by the practical capacity. Conversely with ABC, TDABC does not contain an activity pool in the model (Tse & Gong, 2009). Activities are represented by *time equations*, which are the sum of individual activity times with time drivers. Through a simple time equation it is possible to represent all possible combinations of activities (e.g. different types of products do not necessarily require the same amount of time to be produced). Activity costs are then distributed to cost *objects* by multiplying the cost per time unit of the resources by the estimate of the time required to perform the activities.

3.3.2.1 Time equations

A time equation is a mathematical expression of the time needed to perform activities as a function of several activity time drivers (Hoozée et al., 2009). It implicitly assumes that the duration of an activity is not constant, but a function of the time consumed by the k possible events of an activity and their specific characteristics (i.e., time drivers) (Bruggeman, Everaert, Anderson, & Levant, 2005; Everaert & Bruggeman, 2007). It is represented as follows (Kaplan & Anderson, 2007b):

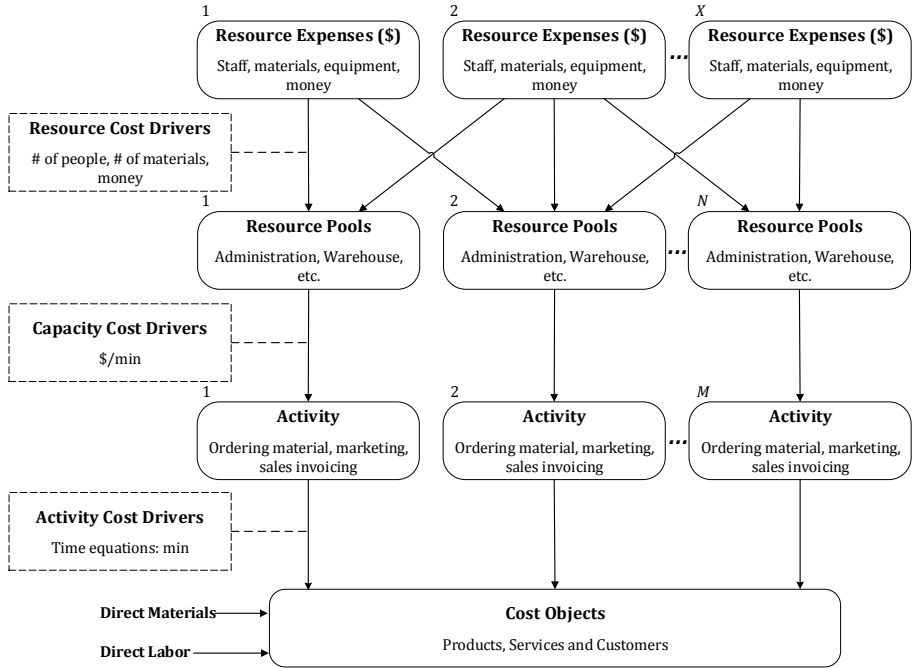


Figure 3.2: TDABC model (based on (Everaert, Bruggeman, Sarens, et al., 2008))

$$T = \text{sum of individual activity times} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_i X_i + \dots + \beta_k X_k$$

With:

T = The time required to perform an activity with k events

β_0 = The basic time to perform the activity (independent of the characteristics of the activity)

β_i = The estimated time for the incremental activity i , with $i = 1, \dots, k$

X_i = The quantity of incremental activity i (transactional data)

k = The number of time drivers taken into account

Time drivers are an essential part in time equations (Everaert & Bruggeman, 2007). They are the characteristics that determine the time needed to perform an activity (Everaert, Bruggeman, De Creus, & Moreels, 2007). Complexity in the process caused by a particular product or order, may add terms but the process is still modeled with only one time equation (Kaplan & Anderson, 2007a).

Time equations can contain three types of variables: continuous, discrete and dummy variables (Everaert & Bruggeman, 2007). Continuous variables are real variables such as the weights of pallets, water temperature or distance in kilometers. Discrete variables are integer variables such as number of orders. These first two types of variables represent standard activities. However, there are certain activities that can influence the formula and that are denoted by indicator variables (Everaert & Bruggeman, 2007). Indicator or dummy variables can only take the value of zero or one (Boolean values) whether the optional activity is or is not used in a particular case. Examples of dummy variables include the type of customer (new or old), the type of order (normal

or rush), the type of shift (morning or evening), etc. The incorporation of these variables in the model simplifies the formulation of the equations (Somapa, Cools, & Dullaert, 2011).

3.3.2.2 *Multiple time drivers and interaction of time drivers*

Multiple time drivers define the time needed to perform an activity and its cost. A time equation provides the ability to include multiple time drivers if an activity is driven by more than one driver (Dalci et al., 2010). It allows to identify and report complex and specialized transactions in a simple way (Everaert et al., 2007). The number of time drivers is unlimited as long as the full complexity is represented in the time equation (Bryon, Everaert, Lauwers, & Van Meensel, 2008). The only restriction is that employees, machinery, etc. performing the tasks must belong to the same resource pool (Everaert & Bruggeman, 2007). According to Varila et al. (2007), the use of multiple variables enables the possibility of collecting more information, the simplification of the estimating process, and the production of a more accurate cost model. It also facilitates a much deeper understanding of the cost behavior of an activity or process. Nevertheless, they also mention that the use of multiple variables will inevitably weaken the traceability of costs.

Another characteristic of time equations is that they might take into account *interactions* between drivers. It is applied if a certain activity depends on the occurrence of other activities, and the activity time is also influenced by the interaction between the two time drivers (Hoozée et al., 2009). It can be represented by the expression below (Everaert & Bruggeman, 2007):

$$T = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

3.3.2.3 *Activity cost*

Once the estimated time for the activity and the unit cost of each resource group are calculated, then the activity cost is computed (Everaert & Bruggeman, 2007). It is represented by the following mathematical expression:

$$\text{Cost of an individual event } k \text{ of activity } j \text{ performed by resource pool } i = t_{j,k} * c_i$$

With:

c_i = The cost per time unit (\$/minute) of resource pool i

$t_{j,k}$ = The time consumed by event k of activity j

Finally, the total cost of a cost object (e.g. process, customer, order, product, service, etc.) over all events of all activities is calculated (Everaert & Bruggeman, 2007). It is done by summing all activity costs and can be represented by the following expression:

$$\text{Total Cost of a Cost object} = \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^l t_{j,k} * c_i$$

With:

c_i = The cost per time unit (\$/minute) of resource pool i

$t_{j,k}$ = The time consumed by event k of activity j

n = The number of resource pools

m = The number of activities

l = The number of times that activity j is performed

3.3.3 TDABC vs. ABC

Adkins (2007) and Yilmaz (2008a) describe ABC as a “push” model, as it starts by estimating the total expenses incurred by the different resources, then by determining the percentage of resources consumed by each product or service, finally by applying this factor to the total cost. Conversely, TDABC is described as a “pull” model because it calculates the total cost by means of the estimation of the unit times required to perform an activity and the cost per time unit of supplying resource capacity. In ABC, the costs of activity-cost pools are apportioned amongst cost objects using activity-cost drivers calculated for each subtask (Kaplan & Cooper, 1998; Tse & Gong, 2009). In TDABC, the costs are allocated to the cost objects on the basis of time units consumed by the activities calculated for the whole department (Kaplan & Anderson, 2004). Unlike ABC, time unit value refers to the time an employee spends doing an activity, and not the percentage of time that it takes to complete one unit of that activity. This type of measurement allows to reduce errors at the moment of calculating the time (Everaert, Bruggeman, Sarens, et al., 2008).

The majority of the differences are based on the weaknesses of ABC (Dejnega, 2011). For instance, Kaplan and Anderson (2004, 2007a) claim that TDABC simplifies the ABC method because of the availability to include multiple time drivers. These time drivers allow to reduce the number of activities and to analyze the costs at the level of the departments or the processes. For instance, the authors present a case study where 1200 activities were reduced in 200 processes. Thus, the model size in TDABC grows linearly with real world complexity (Kaplan & Anderson, 2007b), whereas in ABC growth is exponential to reflect more detail on activities (Kaplan & Anderson, 2004). Likewise, thanks to the use of time equations in TDABC, the high cost and time spent in re-interviewing people every time an activity is changed or updated is reduced because it can be updated based on events rather than by the calendar (Everaert, Bruggeman, & De Creus, 2008).

Barrett (2005) describes the inability of ABC to reveal areas of substantial excess capacity. This is because in ABC, employees tend to ignore their idle time at the moment to estimate the percentage of time spent doing an activity (Yilmaz, 2008a). On the contrary, TDABC is described as a system that automatically exposes differences between the total time needed to perform activities and the total time employees have available. In addition, ABC calculates cost drivers based on the theoretical capacity of the resources supplied while TDABC uses the practical capacity to perform its calculations (Kaplan & Anderson, 2004, 2007a). Eventually, Kaplan and Anderson (2004, 2007a) express the difficulties when integrating ABC with other organizational information systems, while transactional data required for TDABC can easily be obtained from ERP systems, CRM, etc.

3.3.4 Case studies

A Web-based literature research on empirical documents about TDABC applications was conducted in order to identify relevant articles. The data collection was based on materials published in journals, books, Web pages, etc. by means of the search engines: Web of Science, Scholar Google, IDEAS and Google over the period 2004-2012. The search was operated according to the following procedure. First, we searched for relevant articles by combining the terms “time-driven activity-based costing” or “TDABC” with “case study”. Then the existing literature was sorted, summarized and discussed in order to generate a final sample consisting of thirty-six papers. In Table 3.1, the case studies are grouped based on the area that the method is applied and a summary of the main findings is presented for each case study. Strikingly is that a large part of the published case studies on TDABC are case studies in the non-profit sector (health (31%), libraries (8%)). Within the profit sector, a substantial number of case studies have been performed within logistics (31%). Interesting is also to highlight the significant number of case studies conducted in Belgium (25%).

According to Everaert et al. (2008) and Kaplan and Anderson (2007b), TDABC has been successfully implemented for *logistics and service operations*. This is confirmed in our analysis since a large part of the papers (31%) involve the implementation of TDABC in logistics environments. Bruggeman et al. (2005) discuss the application of TDABC in Sanac, a distribution company in Belgium. They define this company as a distributor of plant-care products with a seasonal trend in

Table 3.1: Case studies sorted by area

Paper	Activities	Findings
<i>LOGISTICS</i>		
Bruggeman et al. (2005)	Application of TDABC in a distribution company: Sanac.	Capture the different types of complexities of the logistics transactions through time equations that employ multiple time drivers.
Varila et al. (2007)	The applicability of different drivers for assigning activity costs to products in warehouse logistics environment.	Integration with existing systems to collect information for time equations. Increase the accuracy of accounting by measuring the actual durations. Collect an extensive amount of data in order to increase the understanding of the cost behavior of activities and products.
Oztaysi et al. (2007)	Economic analysis of Radio frequency identification (RFID) technology in courier sector Comparison of a barcode system with a potential RFID system.	Suitable model for investment analysis and comparisons between potential systems. Quantify the performance criteria of rapidity of mail delivery. Allow to perform an economic feasibility study of implementing RFID technology.
Everaert et al. (2008; 2007)	Decision of a distribution company on implementing ABC or TDABC: Sanac.	TDABC drives costs by transactions, enables to reflect complex contingencies in resource-consuming times, assists to improve inefficient processes, and transforms unprofitable customer relationships.
Everaert et al. (2008)	The experiences of the Belgian wholesaler with TDABC in modelling its complex logistics operations.	Provide opportunities to design cost models for complex operations. Capture the variability of the working methods, by including all possible subtasks in the time equation. Provide insight into the causes of excessive logistics and distribution costs.
Gervais et al. (2010)	Longitudinal assessment of the TDABC implementation in the logistic company: SANAC.	Require precise and elaborated analyses that make the starting more lengthy and costly. Use of standard times and costs reduces its complexity. Require regular maintenance. Accuracy is debatable if staff reports their times when it is not possible to observe them directly.
Somapa et al. (2010)	Development of TDABC model in a small-sized road transport and logistics company.	Small-scale firms lack the essential quantitative data to support the buildup of time equations. Difficulty to estimate time in non-continuous activities. Should be implemented a formal time-tracking system. Build and maintain TDABC models with spreadsheets as extensive data is not a problem for small firms.
Hoozée and Bruggeman (2010)	Analysis of four distribution warehouses to examine the role of employee participation and leadership style in the design process of a TDABC system.	Employee participation and leadership style are determinants during the design process of a TDABC system.

Paper	Activities	Findings
Somapa et al. (2011)	Development of TDABC model in a small road transport and logistics company.	Useful for small firms due to the use of simple parameters. Provide accurate cost information in the transport and logistics activities. Capable to calculate the services costs and provide the cause-and-effect relationship between costs and activities. Better resource utilization. Only require the use of spreadsheets to build and maintain TDABC models for small-scale firms.
Ratnatunga et al. (2012)	Implementation of TDABC in the production logistics of a manufacturing company that produces activated carbon in Sri Lanka.	No different to ABC, if standard-activity times are used as cost drivers. Unable to help organizations solving implementation problems in a manner that will not compromise accuracy.
MANUFACTURING		
Bryon et al. (2008)	Supporting the choice of conversion to batch farrowing in pig production.	Guide managers with the decision problem of switching to batch farrowing. Allow them to derive trade-offs between economic and ecological criteria.
Korpunen et al. (2010)	Calculation of the cost of the sawing process.	Assist sawmill management in strategic decision making. Assess the production costs in a sawmill. Enable a sensitivity analysis of production. Accurate, applicable and effective model.
Öker & Adigüzel (2010)	Implementation of TDABC in a manufacturing company.	Provide more relevant information about product profitability and capacity utilization than standard costing.
Stout & Propri (2011)	Implementation of TDABC at a Medium-Sized Electronics Company.	Allow organizations to be accurate with their estimates. Update the model more easily than in ABC models. Capability of implementing TDABC with the support of ERP systems.
Ruiz de Arbuló et al. (2012)	Experience of an auto parts manufacturer who shifted from ABC to TDABC.	Capture the costs of the different products in the product mix. Calculate the cost of a product accurately. Collect data for TDABC seems to be complex. Analyze how capacity is being used (overused or underused).
SERVICES		
Reddy et al. (2011)	Estimating the cost of implementing and managing activities required for digital forensic readiness.	Measure costs at the level of tasks and activities. Less costly and simpler to implement than traditional methods. TDABC is not an automatic fit for any organization. Should be coupled with integrated information systems. Require top management support. Consider potential resistance at the moment of its implementation.
Adeoti & Valverde (2012)	Cost management of Information Technology (IT) Services Operations (Technical Services department).	Identify costly processes which may then allow IT operations managers and supervisors to take critical decisions about cost control, charge-back or costing of services.

Paper	Activities	Findings
<i>HEALTH</i>		
Bank & McIlrath (2009)	Cost estimation of three emergency department (ED) services.	Effective and accurate tool to estimate the true cost of ED services. Help to determine the allocation of ED clinical resources. Help development professional and facility reimbursement strategies with commercial payers.
Demeere et al. (2009)	Analysis of five outpatient clinic's departments: Urology, Gastroenterology, Plastic Surgery, Nose-Throat and Ears, and Dermatology.	Allow managerial recommendations concerning improvement opportunities. Introduce a healthy competition and an open communication between the different departments concerning possible operational improvements. Improve the understanding of the different organizational processes.
Nascimento & Calil (2009b)	Method to create resource consumption profiles for biomedical equipment within medical procedures.	Facilitate the evaluation of equipment used in different conditions. The use of diagrams and tables helps presenting details about the resource consumption structure and the procedure structure itself. Data collection must be driven carefully since it relies on many data estimations.
Nascimento & Calil (2009a)	Resource costs estimation that medical equipment consumes during medical procedures. Cost evaluation of equipment used during an abdominal aortic aneurysm surgery.	Offer a good insight on where the resources are going and the details about the procedure structure itself. Practical tool to evaluate the equipment cost structure of medical procedures. Assess possible resource or practice changes. Flexible to be used in any kind of medical procedure. The quality of results depends on the quality of the data available.
Szucs et al. (2009)	Assessment of the true acquisition cost of erythrocyte concentrate in different European health care systems.	Reveal activities and resources that were excluded in previous accounting attempts. Provide a complete and documented scope respecting the societal perspective.
Bendavid et al. (2010)	Automation of the nursing unit's supply chain with the RFID two-bin replenishment system.	Identify cost centers and assign them to specific activities and processes.
Bendavid et al. (2011)	Evaluation of a RFID-enabled traceability system for a hospital operating room.	Identify cost centers and assign them to specific activities and processes.
Boehler et al. (2011)	Estimation of the cost of changing physical activity behavior.	Facilitate the measurement of resources consumed by individual patients and the allocation of single cost items to each patient.

Paper	Activities	Findings
Box et al. (2012)	True cost estimation of flow cytometry experiments Modelling of new fee schedules	Valuable tool to determine and categorize exact operations costs. Empower the understanding of different types of costs and a better communication to the research organization's leaders. Approach the accounting process to people without background in accounting, business or finance. TDABC cost values should not be used directly to determine service charges.
<i>HOSPITALITY</i>		
Dalci et al. (2010)	Implementation of customer profitability analysis in a Turkish hotel.	Show profitable customer segments which were found unprofitable under the ABC method. Reveal the cost of idle resources.
Hajiha et al. (2011)	Implementation feasibility of TDABC in hospitality industry in Iran.	Provide accurate data on cost and profitability of customers. Distinguish non-added value activities and demonstrate real capacity of each parts of the hotel.
<i>OTHER NONPROFIT SERVICES</i>		
Pernot et al. (2007)	Costing analysis of inter-library loans	Visualize activities that consume the largest amount of time. Enable to disaggregate per-transaction costs. Visualize the true cost of different activities. Allow managerial recommendations concerning amelioration opportunities.
Stouthuysen et al. (2010)	Implementation of a TDABC model in a library acquisition process	Identify several factors that drive the cost of the acquisition process of library items. TDABC is well suited for a library setting, involving many activities with complex time drivers. Create more visibility to acquisition process efficiencies and capacity utilization.
Ratnatunga & Waldmann (2010)	Understanding of university research activities through an accounting technique.	Provide accurate information in "research only" departments and institutes. Inappropriate for determining the indirect research costs of teaching departments.
Everaert et al. (2012)	Introduction of TDABC in a university restaurant in order to allocate operating expenses to cost objects.	Time equations facilitate a detailed understanding of the work activities. The more detailed accounting system offers insights into bottlenecks and capacity utilization of employees.
Siguenza-Guzman et al. (2013)	Application of TDABC to support loan and return processes	Better understanding of the costs' origin due to the disaggregated values per activity. An improved alternative evaluation to compare different scenarios. An enhanced communication to analyze the cause of specific problems with stakeholders that can easily understand the methodology. Adaptability when for instance it is required to switch resources in busy periods.

the sales and diversity in resource consumptions. In this case study, TDABC is described as a useful cost model to be applied in small and medium-sized enterprises as well as in environments with complex activities such as logistics, hospitals, distribution and servicing companies in general. The authors conclude that TDABC is able to capture the full complexity of the logistics transactions thanks to the use of time equations and multiple time drivers. Everaert et al. (2007; 2008a) describe the decision making of the same company on whether to proceed with the implementation of an ABC system or switch to TDABC. They report that ABC failed to capture the complexity of the company, whereas TDABC could drive costs by transactions and enabled to reflect complex contingencies in resource-consumption times. The authors conclude that TDABC assists managers to improve inefficient processes and to transforms unprofitable customer relationships into profitable relationships. Everaert et al. (2008b) report the two-year experience of implementing TDABC in a Belgian wholesaler. The authors conclude that TDABC is able to trace the full complexity of the logistics operations, capture the variability of the working methods by including all possible subtasks in the time equation, and provide insight into the causes of excessive costs. Eventually, Gervais et al. (2010) perform a longitudinal assessment of the four-year implementation of TDABC in the same company. They conclude that TDABC offers a partial solution to the weaknesses criticized on ABC especially regarding the cost and complexity of implementing and maintaining it as the data gathering process is still substantial. Similarly, Somapa et al. (2010, 2011) report the development of TDABC model in a small-sized road transport and logistics company in Thailand and they also describe some difficulties associated with the study of a small scale operation such as insufficient data to support the time equations and the lack of a time-tracking system (e.g. log book for a manual record). On the other hand, they conclude that the data needed in small-scale firms is not very extensive so that they were able to build and maintain TDABC models with spreadsheets. In addition, they report several benefits of using TDABC such as the application of detailed costs based on the service routes and the different destination types, the revelation of loss-generating routes, the identification of the cause of loss and the utilization of company resources. Varila et al. (2007) use TDABC to examine the applicability of different drivers for assigning activity costs to products in warehouse logistics environment. The authors state that by integrating TDABC with existing systems to collect information for time equations, it enhances the accuracy of accounting. They recommend collecting an extensive amount of data in order to increase the understanding of the cost behavior of activities and products. Oztaysi et al. (2007) apply TDABC to economically justify the use of RFID technology in the courier sector. They compare the current barcode system with a potential RFID system. They conclude that TDABC is a suitable model for investment analysis and comparison between potential systems. Hoozée and Bruggeman (2010) conduct a case study in four distribution warehouses at a Belgian division of a company to illustrate that employee participation and leadership style are key factors at the moment of implementing TDABC. Finally, the most recent case study in logistics presented by Ratnatunga et al. (2012), makes a comparison of TDABC with ABC by implementing both methodologies in the production logistics of a manufacturing company that produces activated carbon in Sri Lanka. The authors conclude that TDABC seems no different to ABC if standard-activity times are used as cost drivers. They state that TDABC has similar implementation complexities to ABC.

TDABC has been successfully implemented in several *other sorts of companies* (Yilmaz, 2008a). For instance, Bryon et al. (2008) present an implementation of TDABC in the farming industry. The model is applied for a specific decision problem and adapted for multi-criteria evaluation to analyze whether pig farmers should switch from traditional pig production to batch farrowing. The authors describe TDABC as a useful system to quantify both the economic and ecological impacts of strategic decisions. Korpunen et al. (2010) utilize TDABC to calculate the production cost of the sawing process. The authors conclude that TDABC is a promising method to assist sawmill's managers in strategic decision making because it allows to assess the production costs and to perform a sensitivity analysis of production. Öker and Adigüzel (2010) describe the implementation of TDABC in a manufacturing company to conclude that TDABC provides relevant information about product profitability and capacity utilization. Stout & Propri (2011) implement TDABC at a medium-sized electronics company to demonstrate both the potential power of TDABC and the important role of ERP systems in implementing TDABC systems. They conclude that TDABC allows organizations to be accurate with their estimates, to update the model more easily than ABC models and to allocate support cost to products and customers thanks to the integration of ERP

systems. In a recent study presented by Ruiz de Arbulo (2012), TDABC is described as a system that can capture the costs of the different products, can accurately calculate the cost of a product and can help to analyze the capacity utilization. The authors describe the experience of an auto parts manufacturer who shifted from ABC to TDABC. Despite the number of advantages stated, the authors indicate that data collection for TDABC is complex.

Szychta (2010) offers a synthetic description of the validity of TDABC applications in *service companies*. The author states that this model is suitable for service activities because they are primarily measured on the basis of the labor time used for performing a given activity. This statement is confirmed in our analysis because the majority of the papers implement TDABC in the service sector (e.g. health, hospitality, libraries, etc.). For instance, Reddy et al. (2011) estimate the cost of implementing and managing activities required for digital forensic readiness. The authors conclude that TDABC can be used to determine the costs of DRF because it allows to measure cost at the level of tasks and activities and is less costly and simpler to implement than ABC. In order to successfully implement TDABC, the authors consider that there are certain factors to be taken into account such as the coupling TDABC with integrated information systems, the top management support or the incompatible organizational culture. Adeoti and Valverde (2012) apply TDABC to the Technical Services department of Information Technology (IT) Services Operations. The authors conclude that TDABC is an effective tool to identify costly processes in IT operations and thus to take critical decisions about cost control, charge-back or costing of services.

Within the TDABC literature, a lot of research attention has been devoted to applying TDABC in *healthcare sector* (31%). In this regard, Bank and McIlratha (2009) use data from a high volume pediatric emergency department (ED) to estimate the costs of providing resources and to apply them into three specific clinical scenarios common in any ED service. They consider TDABC to be useful to ED directors in helping to determine the allocation of clinical resources. Demeere et al. (2009) explain how to implement a TDABC model for five outpatient clinic departments (Urology, Gastroenterology, Plastic Surgery, Nose-Throat and Ears and Dermatology). The authors conclude that TDABC offers some additional benefits to those offered by ABC, such as faster model adaptability, a simpler set-up and a higher reflection of the complexity of the real-world operations. They also suggest that TDABC is a useful system to understand the different organizational processes in a healthcare environment. Nascimento and Calil (2009a, 2009b) present two case studies using TDABC to estimate the resource costs that equipment consumes during medical procedures. The authors conclude that TDABC is a flexible system that facilitates the evaluation of equipment used in different conditions. They recommend managing the data collection carefully since the quality of results depends on the quality of the data available. Szucs et al. (2009) describe TDABC to assess the true acquisition cost of blood products in different European health care systems. According to the authors, TDABC is a complete and transparent cost allocation methodology that provides a detailed process description of activities and resources excluded in previous accounting systems. Bendavid et al. (2010, 2011) utilize TDABC to automate and evaluate RFID systems in a hospital operation room and a nursing unit's supply chain. They select TDABC because of the need to identify cost centers and assign them to specific activities and processes. Boehler et al. (2011) use TDABC to estimate the cost of changing physical activity behavior. The authors select TDABC as it facilitates the measurement of resources consumed by individual patients and the allocation of single cost items to each patient. The most recent case study in the health care sector presented by Box et al. (2012) estimates the cost of flow cytometry experiments. The authors describe TDABC as a valuable tool to determine the true costs of flow cytometry experiments and to determine where significant discrepancies in cost recovery occur. TDABC also helps to model new fee schedules with the goal of more accurately reflecting resource usage in fees. In addition, TDABC is an approach that is easily understandable to people without a background in accounting, business or finance. Despite these advantages, the authors also state that for flow cytometry experiments, TDABC should not be used directly to determine user charges (i.e., the precise cost for a service as determined by TDABC should not be equal to the charge for that service). An important reason is that service charges should be predictable: a rate schedule should be easily recalled and/or predictable so that the financial expectations for users are clear. Another reason is that sometimes it may be useful to intentionally decouple actual costs from service

charges to guide the use (and therefore capacity) of resources. For example, a reduced charge rate for after-hour use of an instrument can increase the instrument's overall used capacity.

Two papers deal with the application of TDABC in the *hospitality industry*. Dalci et al. (2010) describe the implementation of customer profitability analysis (CPA) using TDABC in a hotel. CPA describes the process of assigning costs and obtaining revenues to customer segments or individual customers' accounts in order to calculate the profitability of the segments or accounts (Raaij, 2005). The case study reveals the cost of idle resources devoted to the front desk, housekeeping, food preparation, and marketing activities. They conclude that TDABC provides valuable information to support managerial decision making in a hotel. However, they suggest replicating the analysis to other similar scenarios to see if the results could be generalized. Hajiha & Alishah (2011) present positive results of implementing TDABC in the hospitality industry in Iran. The authors conclude that TDABC provides more proper data on cost and profitability of customers than traditional costing systems. They also state the proposed model distinguishes non-added value activities and demonstrates real capacity of each part of the hotel.

Five papers focus on the application of TDABC on *other nonprofit services*, of which three papers specially focus on the implementation of TDABC in libraries. The first TDABC approach in libraries given by Pernot et al. (2007) calculates the costs of the inter-library loan service. They argue that TDABC can improve the cost management of all library services because it enables to disaggregate "per-transaction costs" which allows taking appropriate actions to improve time consuming activities. Stouthuysen et al. (2010) describe TDABC as a useful system for small to medium-sized academic libraries. The study is focused on the acquisition process. The authors claim that TDABC assists managers in visualizing the acquisition process efficiencies and capacity utilization, leading to potential cost efficiencies. They also state that TDABC can be applied to complex or digitalized acquisition environments. In a recent study, Siguenza-Guzman et al. (2013) present a case study on the loan and return processes. They compare the costs of some specific activities performed by staff or machines, concluding that the usage of robots is well-justified to automate repetitive processes. The authors demonstrate that TDABC is applicable to large library activities, but that involvement of library staff during the TDABC implementation is crucial. They also conclude that TDABC leads to an effective process analysis and better decision-making by librarians and library administrators. Ratnatunga and Waldmann (2010) determine the costs of Australian Competitive Grant (ACG) research projects with the objective of ensuring the full funding of these projects by the government. The authors consider TDABC inappropriate for teaching and research departments since accurate estimations could be obtained from other sort of methods such as studying the workload allocation, and conducting direct interviews of the staff undertaking research on ACG or other externally funded grants. On the contrary, the use of TDABC for "research only" departments is highly recommended as it is possible to obtain accurate estimation based on 'in-situ' observations, face-to-face interviews and the study of comparative information. Eventually, Everaert et al. (2012) utilize TDABC to calculate the costs of meals offered in a university restaurant. They use a "V-structure" to create a link between different levels of costs objects. In order to determine the cost per meal, the food cost of each meal component is directly assigned. For all other costs such as preparing, serving and serving the meal, TDABC is employed. The authors describe TDABC as a costing technique that offers the benefits of ABC with less administration costs. They support the idea that time equations of the meal process allow a detailed understanding of working activities (e.g. a lunch or dinner session consisting of different meals, and meals composed of several meal components). In addition, thanks to the time equations, operational improvements can be identified such as the recommendation of offering less profitable meals on calmer days, while offering more profitable meals on busy days.

3.4 Benefits and challenges

3.4.1 Benefits

Based on the different case studies, we can expose a number of benefits.

1) *Simplicity*. According to Kaplan and Anderson (2007b), this is the most important attribute of TDABC. It is confirmed by Somapa et al. (2011) who recommend TDABC because of the use of two simple parameters: the cost per time unit of the activity, and the time required to perform an activity. Thanks to this simplicity, TDABC allows to approach the accounting process to people without experience in accounting, business or finance (Box et al., 2012; Siguenza-Guzman et al., 2013). TDABC also allows to improve the understanding of the different organizational processes through the lens of an accounting technique (Box et al., 2012; Demeere et al., 2009; Ratnatunga & Waldmann, 2010).

2) *Complex Operations*. TDABC allows to design cost models for complex operations thanks to the use of multiple time drivers (Boehler et al., 2011; Everaert, Bruggeman, Sarens, et al., 2008; Nascimento & Calil, 2009b). TDABC captures the variability of the activities by including all possible subtasks in the time equation (Everaert et al., 2007; Everaert, Bruggeman, Sarens, et al., 2008; Stouthuysen et al., 2010). In turn, time equations can include multiple time drivers without expanding the number of activities. By using multiple time drivers, TDABC allows to disaggregate per-transaction costs and thereby identify processes that are costly, wasteful and inefficient (Everaert et al., 2012; Kaplan & Anderson, 2007b; Pernot et al., 2007; Reddy et al., 2011).

3) *Capacity utilization*. TDABC allows to have a good estimation of resource consumption and capacity utilization (Bank & McIlrath, 2009; Nascimento & Calil, 2009a; Öker & Adigüzel, 2010; Stouthuysen et al., 2010). According to Szucs et al. (2009) and Dalci et al. (2010), TDABC reveals activities, resources and costs that were excluded in previous accounting attempts. TDABC provides insight into the causes of excessive time or costs occupied by the resources (Everaert, Bruggeman, Sarens, et al., 2008). Managers can review the time and cost of the unused or overused capacity and contemplate actions to improve them (Demeere et al., 2009; Kaplan & Anderson, 2007a; Ruiz de Arbulo et al., 2012). They may also reserve resources for future growth instead of reducing currently unused capacity (Kaplan & Anderson, 2007a).

4) *Versatility and modularity*. According to Stout & Propri (2011), TDABC can be updated more easily than ABC models. Kaplan and Anderson (2007a) state that managers do not have to re-interview personnel when more activities are added to a process. For cost drivers, there are two factors that cause a change: 1) changes in the costs of resources supplied affecting the capacity; 2) modified or updated processes such as new or redesigned processes, products, channels, etc. (Everaert & Bruggeman, 2007; Kaplan & Anderson, 2007a). It can be updated based on events rather than by the calendar (Kaplan & Anderson, 2007a). Likewise, by implementing TDABC with the support of existing systems (ERP systems, CRM, etc.), the system allows for an easy updating as well as a greater accuracy (Kaplan & Anderson, 2007b; Ruiz de Arbulo et al., 2012; Stout & Propri, 2011; Varila et al., 2007). The major benefit occurs when companies link their own information systems to the TDABC system (Hudig, 2007; Reddy et al., 2011). In the case of small-scale environments, TDABC can be built and maintained with relatively simple spreadsheets (Somapa et al., 2010, 2011).

5) *Process simulation*. TDABC can be used in a predictable manner. Managers can modify the behavior of their customers by simulating the future through the use of dynamic what-if analysis (Kaplan & Anderson, 2007a). They can also establish future investment decisions due to the possibility to determine the impact of changes in terms of cost, profit, capacity and time (Acorn Systems, 2007; Hudig, 2007).

3.4.2 Challenges

Despite the advantages presented by its authors and supporters, criticisms have been made about TDABC.

1) *Measurement error*. Cardinaels and Labro (2008) report via an experimental analysis that the employees' estimation may not be as accurate as the authors proclaim. A significant degree of subjectivity is still present (Barrett, 2005). Although Kaplan and Anderson (2007b) state that time

consumption data can be estimated or observed directly, it still requires a series of interviews with employees. These interviews may adversely affect the work performed by people in charge of enforcing TDABC approach. Hoozée and Bruggeman (2010) show an example where operational employees feel that by implementing TDABC they are being controlled. Gervais et al. (2010) also present a case study in which certain employees and senior staff members are strongly opposed to state their working time precisely. Reddy et al. (2011) recommend to consider potential resistance at the moment of implementing TDABC models. Cardinaels and Labro (2008) also find a strong overestimation bias when employees provide their time estimates in minutes. Indeed, if the minutes-based model is compared to the percentage-based model (ABC approach), outcomes are definitely more accurate. This is so because few employees tend to report the percentage of their idle time. However, the authors state that more than 77% of their participants consistently overestimated the time spent on all activities. Accuracy is debatable if staff reports their times when it is not possible to observe them directly (Gervais et al., 2010).

2) *Data*. Varila et al. (2007) emphasize the necessity of a considerable amount of data to estimate satisfactorily time equations. This situation is confirmed by Ruiz de Arbuló et al. (2012) and Nascimento & Calil (Nascimento & Calil, 2009a, 2009b) which indicate that data collection for TDABC seems to be complex and should be managed carefully since the quality of results depends on the quality of the data available. Gervais et al. (2010) claim that TDABC requires a precise and elaborated analysis, making the starting more lengthy and costly. They also stress the necessity of a regular maintenance over time (Gervais, Levant, & Ducrocq, 2009). In addition, for Barrett (2005), TDABC depends on robust and reliable data to deliver an acceptable level of accuracy. If the data comes from automated software and is regularly updated, then the results will probably be accurate. However, if the information is out of date, or if it is based on estimates, the resulting cost information may include substantial errors (Sherratt, 2005). Barret states that for the estimation of only one activity in minutes, a difference of seconds may not seem important. But, when this time in minutes is multiplied by the number of times an activity is performed in a certain period, the difference becomes significant. The author also objects to the accuracy of the costing process results because of the assumption that practical capacity can be calculated as a percentage of the theoretical capacity.

3) *Dedicated to homogeneous and repetitive activities*. Sherratt (2005) states that TDABC is limited to predetermined routines and activities. Barrett (2005) considers TDABC to be simple for a department that performs a single activity, because the total costs of the direct and indirect resources can be divided by the available resource. However, most departments perform more activities that consume direct and indirect resources in different proportions (Barrett, 2005). Wegmann and Nozile (2010; 2009) claim that TDABC is only useful for standard processes such as chain management supplying, some standardized production processes and consulting activities, call centers, hospitals, etc. Ratnatunga & Waldmann (2010) consider the use of TDABC inappropriate to determine the indirect research cost for departments that combine teaching and research activities since accurate estimations can be obtained from other type of methods. Cardinaels and Labro (2008) also state that for incoherent tasks such as research and development process, marketing, legal or complex productions, some mistakes are possible. Incoherent tasks are activities addressed on the basis of first come, first served, and not in a structured and systematic sequence. In these cases, authors predict less accuracy in time perception when events are presented incoherently rather than coherently. Sherratt (2005) also considers that certain areas of IT and marketing do not perform repetitive and homogeneous activities to be clocked reliably. For these activities, the author still considers ABC as an alternative methodology to be used (Barrett, 2005). This situation is confirmed by Hoozée et al. (2010), who also affirm that ABC is more accurate than TDABC in those types of cases. Through a simulation analysis, the authors compare the overall accuracy of ABC and TDABC in complex and dynamic environments. They identify that when diversity of productive work is low, TDABC tends to be more accurate, especially at higher levels of unused capacity. Conversely, when the diversity of productive work is high, ABC is the best option, especially at lower levels of unused capacity.

3.5 Opportunities

There are several additional opportunities for TDABC such as:

1) *Simulation*. McGowan (2009) states that TDABC may easily use simulation modeling to analyze how to optimize the resources since information is entirely composed of real values. It allows implementing different scenarios in order to identify opportunities for resource management. Moreover, the simulation model can be used for capacity planning to highlight resources gaps and spare capacity (Everaert, Bruggeman, Sarens, et al., 2008). Managers can easily update their simulation model to reflect changes in the operating environment and to measure improvement in efficiencies and costs.

2) *Benchmarking*. With TDABC, companies can compare their processes among other companies because most of them are common across multiple industries (Kaplan & Anderson, 2007b). It is also possible to compare time equations and costs within different company locations (warehouses). Anderson (2006) presents how three companies in different industries use TDABC in a benchmarking way. For this author, TDABC does not replace traditional benchmarking methodologies; it enhances them. This is so because, unlike traditional benchmarking that only reports macro results, TDABC isolates process differences to uncover root causes. Everaert et al. (2008) present a case study where an internal benchmarking was performed in four warehouses. They describe that time estimations were different among all the warehouses. In some cases, this is because of the different distances that trucks must travel; and for other subtasks, because some warehouses worked more efficiently than others.

3) *Complementary Information Systems*. A fully automated accounting mode is not yet usual. Nevertheless, there are a variety of technologies that help to collect data from processes with minimum manual efforts such as bar codes, RFID technology and time sheets (Bahr & Lim, 2010; Oztaysi et al., 2007; Varila et al., 2007). This is especially useful in a logistics environment where there are mostly repetitive processes. Moreover, since these technologies allow improving data accuracy by reducing the number of human errors and because they allow a relatively rapid data collection, they may provide an answer for TDABC for non-routine tasks. Varila et al. (2007) also recommend to estimate the time on the bases of parameters through statistical tools such as multiple regression analysis. Neural networks are also another kind of flexible tools to estimate costs, although they are a "black box" which gives no chance to analyze the outcomes (Verlinden, Dufloy, Collin, & Catrysse, 2008).

4) *Balanced Scorecard (BSC)*. According to Yilmaz (2008a), TDABC can be used as a basis for a balanced scorecard. The BSC allows organizations to implement a strategy rapidly and effectively by integrating the measurement system with the management system. TDABC facilitates translating strategy into performance measures and provides actionable performance measures for the BSC (Yilmaz, 2008a). Yilmaz (2008b) and Ayvaz et al. (2011) analyze the relation between BSC and the costing systems ABC and TDABC respectively. They mention four existing links: 1) *Operational Connection*, where the outputs of ABC such as costs, quality, time and innovations are usually excellent inputs to a BSC by defining the performance of any process. 2) *Customer Profitability Connection*, through TDABC because of the ability to accurately decompose the aggregate marketing, distribution, technical, service, and administrative costs into the cost of serving individual customers. 3) *Financial Connection* because BSC helps to identify the strategic initiatives and resource requirements that enable companies gaining sustainable competitive power in the long term. Resources for these initiatives are assigned in the annual spending budget. 4) *Analytic Hierarchy Process*, which allows decision makers to model a complex problem in a hierarchical structure.

5) *Total Quality Management (TQM)*. TQM means excellent quality of products and services and its objective is to meet customer requirements through the involvement of all the employees (Novičević & Ljilja, 1999). It stresses the need to manage the activities and processes in a

continuous improvement framework. In the case of traditional costing systems, they cannot be adapted to the TQM philosophy because they are directed to the product and not to the process. In Novičević and Ljilja (1999), ABC is analyzed to deal with this problem since it provides information on costs and also information on processes. It helps managers in realization of a cost reduction program by the premise that certain costs can be eliminated, and that action does not make product quality an inferior one. It can be done with the objective to eliminate activities that do not add value to the product. According to Novičević and Antić (1999), ABC is totally compatible with TQM philosophy.

3.6 Conclusions

This document presents a comprehensive literature review on Time-Driven Activity-Based Costing, with a special focus on the case studies published over the period 2004-2012. Thirty-six papers are analyzed and classified along application themes such as logistics, manufacturing, services, health, hospitality and other nonprofit services. Based on the analysis of the selected literature, we conclude that TDABC is highly recommended for repetitive activities. However, the current research is less clear about the advantages of TDABC for non-routine tasks. Technologies such as RFID, bar codes or existing information from time sheets may provide the necessary data required in these cases. Nevertheless, future research is needed to identify whether these data sources may be helpful for non-routine tasks. Comparing TDABC to the traditional ABC costing, TDABC offers several advantages, even if it does not dramatically simplify specific processes. We agree with Adkins (2007) that TDABC should be considered as a complement to the traditional ABC model rather than as a replacement of it.

The results show that in practice TDABC provides most of the advantages its authors claim. Nevertheless, despite advantages, the main remarks made by other authors, which require special attention are: 1) provision of a partial solution to the ABC failings, 2) difficulty to measure the times, the homogeneity and their maintenance over time, 3) degree of subjectivity still present in the model, 4) biased overestimation when employees provide their time estimates in minutes, 5) necessity of a considerable amount of data to estimate satisfactorily time equations, 6) dependence of robust and reliable data to deliver an acceptable level of accuracy, 7) necessity of a regular maintenance with a minimum of required knowledge, 8) limitation of the model to predetermined routines and activities (e.g. repetitive processes), 9) difficulties of estimating time for non-continuous or unpredictable activities.

The studies on the implementation, as well as the criticisms on TDABC are, in most of the cases, written by its creators and not by independent researchers. This can certainly bias the evaluation of the TDABC methodology. Future research is needed by operational case studies in specific areas such as the public services, and in activities that follow an unstructured and non-systematic sequence.

References

- Acorn Systems. (2007). Higher Profits, Increased Efficiency With Time-Driven Activity-Based Costing — Retail Solutions Online Interview with Steven Anderson. Retail Solutions Online, 5.
- Adeoti, A., & Valverde, R. (2012). A Time-Driven Activity Cost Approach for the Reduction of Cost of IT Services: A Case Study in the Internet Service Industry. AMCIS 2012 Proceedings, (Paper 4). Retrieved from <http://aisel.aisnet.org/amcis2012/proceedings/AccountingInformationSystems/4>
- Adkins, T. (2007). Five Myths About Time-Driven Activity-Based Costing: Sorting through the facts about traditional and time-driven costing methodologies (White paper) (p. 10). SAS. Retrieved from http://www.sas.com/resources/whitepaper/wp_4034.pdf

- Anderson, S. R. (2006). Maximize Benchmarking with Time-Driven ABC: New Techniques that Change How We Measure Performance. *Acorn Systems*, 1–11.
- Ayvaz, E., & Pehlivanli, D. (2011). The Use of Time Driven Activity Based Costing and Analytic Hierarchy Process Method in the Balanced Scorecard Implementation. *International Journal of Business and Management*, 6(3), 146–158.
- Bahr, W., & Lim, M. K. (2010). Maximising the RFID benefits at the tyre distribution centre. In *Proceedings of the 12th International MITIP Conference The Modern Information Technology in the Innovation Processes of the Industrial Enterprises*. Presented at the 12th International MITIP Conference, Denmark. Retrieved from http://eprints.aston.ac.uk/15223/1/Bahr_Lim_Maximising_the_RFID_benefits_at_the_tyre_distribution_centre.pdf
- Bank, D. E., & McIlrath, T. (2009). 13: Utilizing Time-Driven Activity-Based Costing in the Emergency Department. *Annals of Emergency Medicine*, 54(3, Supplement), S5. doi:10.1016/j.annemergmed.2009.06.033
- Barrett, R. (2005). Time-Driven Costing: The Bottom Line on the New ABC. *Business Performance Management Magazine*, 11(Supplement), 35–39.
- Bendavid, Y., Boeck, H., & Philippe, R. (2010). Redesigning the replenishment process of medical supplies in hospitals with RFID. *Business Process Management Journal*, 16(6), 991–1013. doi:10.1108/14637151011093035
- Bendavid, Y., Boeck, H., & Philippe, R. (2011). RFID-Enabled Traceability System for Consignment and High Value Products: A Case Study in the Healthcare Sector. *Journal of Medical Systems*, 1–17. doi:10.1007/s10916-011-9804-0
- Bjørnenak, T., & Mitchell, F. (2002). The development of activity-based costing journal literature, 1987–2000. *European Accounting Review*, 11(3), 481–508. doi:10.1080/0963818022000001019
- Boehler, C. E. H., Milton, K. E., Bull, F. C., & Fox-Rushby, J. A. (2011). The cost of changing physical activity behaviour: evidence from a “physical activity pathway” in the primary care setting. *Bmc Public Health*, 11(370), 12. doi:10.1186/1471-2458-11-370
- Box, A. C., Park, J., Semerad, C. L., Konnesky, J., & Haug, J. S. (2012). Cost accounting method for cytometry facilities. *Cytometry Part A*, 81A(6), 439–444. doi:10.1002/cyto.a.22052
- Bruggeman, W., Everaert, P., Anderson, S. R., & Levant, Y. (2005). Modeling Logistics Costs using Time-Driven ABC: A Case in a Distribution Company (Working Paper No. 2005/332) (p. 47). Belgium: Faculty of Economics and Business Administration, Ghent University. Retrieved from <http://cosmic.rzr.uni-hamburg.de/webcat/hwwa/edok05/be193g/wp05-332.pdf>
- Bryon, K., Everaert, P., Lauwers, L., & Van Meensel, J. (2008). Time-driven activity-based costing for supporting sustainability decisions in pig production (p. 29). Presented at the Corporate Responsibility Research Congress, Belfast (UK): CRR Conference. Retrieved from <http://crrconference.org/downloads/lauwers.pdf>
- Cardinaels, E., & Labro, E. (2008). On the Determinants of Measurement Error in Time-Driven Costing. *The Accounting Review*, 83(3), 735–756. doi:10.2308/accr.2008.83.3.735
- Clarke, P. J., Hill, N. T., & Stevens, K. (1999). Activity-Based Costing in Ireland: Barriers to, and opportunities for, change. *Critical Perspectives on Accounting*, 10(4), 443–468. doi:10.1006/cpac.1997.0197
- Colwyn Jones, T., & Dugdale, D. (2002). The ABC bandwagon and the juggernaut of modernity. *Accounting, Organizations and Society*, 27(1–2), 121–163. doi:10.1016/S0361-3682(01)00035-6
- Cooper, R., & Kaplan, R. S. (1988). Measure Costs Right: Make the Right Decision. *Harvard Business Review*, 66(5), 96–103.
- Cooper, R., & Kaplan, R. S. (1991). Profit Priorities from Activity-Based Costing. *Harvard Business Review*, 69(3), 130–135.
- Dalci, I., Tanis, V., & Kosan, L. (2010). Customer profitability analysis with time-driven activity-based costing: a case study in a hotel. *International Journal of Contemporary Hospitality Management*, 22(5), 609–637. doi:10.1108/09596111011053774
- Dejnega, O. (2011). Method Time Driven Activity Based Costing - Literature Review. *Journal of Applied Economic Sciences*, 6(1(15)), 7–15.

- Demeere, N., Stouthuysen, K., & Roodhooft, F. (2009). Time-driven activity-based costing in an outpatient clinic environment: Development, relevance and managerial impact. *Health Policy*, 92(2–3), 296–304. doi:10.1016/j.healthpol.2009.05.003
- Ellis-Newman, J., & Robinson, P. (1998). The cost of library services: Activity-based costing in an Australian academic library. *The Journal of Academic Librarianship*, 24(5), 373–379. doi:10.1016/S0099-1333(98)90074-X
- Everaert, P., & Bruggeman, W. (2007). Time-driven activity-based costing : exploring the underlying model. *Journal of cost management*, 21(2), 16–20.
- Everaert, P., Bruggeman, W., & De Creus, G. (2008). Sanac Inc.: From ABC to time-driven ABC (TDABC) – An instructional case. *Journal of Accounting Education*, 26(3), 118–154. doi:10.1016/j.jaccedu.2008.03.001
- Everaert, P., Bruggeman, W., De Creus, G., & Moreels, K. (2007). Chapter Nine. SANAC Logistics. Time Equations to Capture Complexity in Logistics Processes. In *Time-Driven Activity-Based Costing: A simpler and more powerful path to higher profits* (pp. 165–178). Harvard Business Press.
- Everaert, P., Bruggeman, W., Sarens, G., Anderson, S. R., & Levant, Y. (2008). Cost modeling in logistics using time-driven ABC: Experiences from a wholesaler. *International Journal of Physical Distribution & Logistics Management*, 38(3), 172–191. doi:10.1108/09600030810866977
- Everaert, P., Cleuren, G., & Hoozée, S. (2012). Using time-driven ABC to identify operational improvements: A case study in a university restaurant. *Cost Management*, 26(2), 41–48.
- Gervais, M., Levant, Y., & Ducrocq, C. (2009). Time Driven Activity Based Costing: New Wine, or Just New Bottles? In *Proceeding of the 32nd Annual congress of the European Accounting Association* (p. 31). Presented at the 32nd Annual congress of the European Accounting Association. doi:10.1016/halshs.archives-ouvertes.fr:halshs-00391265
- Gervais, M., Levant, Y., & Ducrocq, C. (2010). Time-Driven Activity-Based Costing (TDABC): An Initial Appraisal through a Longitudinal Case Study. *JAMAR (The Journal of Applied Management Accounting Research)*, 8(2), 20 p.
- Gosselin, M. (2006). A Review of Activity-Based Costing: Technique, Implementation, and Consequences. In *Handbooks of Management Accounting Research* (Vol. Volume 2, pp. 641–671). Elsevier. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1751324306020086>
- Gunasekaran, A., & Sarhadi, M. (1998). Implementation of activity-based costing in manufacturing. *International Journal of Production Economics*, 56–57(0), 231–242. doi:10.1016/S0925-5273(97)00139-4
- Hajiha, Z., & Alishah, S. S. (2011). Implementation of Time-Driven Activity-Based Costing System and Customer Profitability Analysis in the Hospitality Industry: Evidence from Iran. *Economics and Finance Review*, 1(8), 57–67.
- Hoozée, S., & Bruggeman, W. (2010). Identifying operational improvements during the design process of a time-driven ABC system: The role of collective worker participation and leadership style. *Management Accounting Research*, 21(3), 185–198. doi:10.1016/j.mar.2010.01.003
- Hoozée, S., Vanhoucke, M., & Bruggeman, W. (2010). Comparing the accuracy of ABC and time-driven ABC in complex and dynamic environments: a simulation analysis (Working Papers of Faculty of Economics and Business Administration, Ghent University, Belgium No. 2010/645). Ghent University, Faculty of Economics and Business Administration. Retrieved from <http://ideas.repec.org/p/rug/rugwps/10-645.html>
- Hoozée, S., Vermeire, L., & Bruggeman, W. (2009). A risk analysis approach for time equation-based costing (Working Paper No. 2009/556) (p. 47). Belgium: Faculty of Economics and Business Administration. Ghent University. Retrieved from http://feb1.ugent.be/nl/Ondz/WP/Papers/wp_09_556.pdf
- Hudig, J.-W. (2007). Robert S. Kaplan: "Better information with less effort". *Fi Next - Wrap up*, pp. 5–6.
- Innes, J., Mitchell, F., & Sinclair, D. (2000). Activity-based costing in the U.K.'s largest companies: a comparison of 1994 and 1999 survey results. *Management Accounting Research*, 11(3), 349–362. doi:10.1006/mare.2000.0135
- Kaplan, R. S., & Anderson, S. R. (2004). Time-Driven Activity-Based Costing - Tool Kit. *Harvard Business Review*, (82), 131–138.

- Kaplan, R. S., & Anderson, S. R. (2007a). The innovation of time-driven activity-based costing. *Journal of cost management*, 21(2), 5–15.
- Kaplan, R. S., & Anderson, S. R. (2007b). *Time-driven activity-based costing: a simpler and more powerful path to higher profits*. Harvard Business Press.
- Kaplan, R. S., & Cooper, R. (1998). *Cost & effect: using integrated cost systems to drive profitability and performance*. Harvard Business Press.
- Korpunen, H., Mochan, S., & Uusitalo, J. (2010). An activity-based costing method for sawmilling. *Forest Products Journal*, 60(5), 420–431.
- Linn, M. (2007). Budget systems used in allocating resources to libraries. *Bottom Line: Managing Library Finances*, The, 20(1), 20–29. doi:10.1108/08880450710747425
- McGowan, C. (2009). Time-Driven Activity-Based Costing A New Way To Drive profitability. *Accountancy Ireland*, 41(6), 60–61.
- Michalska, J., & Szwieczek, D. (2007). The improvement of the quality management by the activity-based costing. *Journal of Achievements in Materials and Manufacturing Engineering*, 21(1), 91–94.
- Nascimento, L. N., & Calil, S. J. (2009a). A Method to Create Resource Consumption Profiles for Biomedical Equipment. In O. Dossel & W. C. Schlegel (Eds.), *World Congress on Medical Physics and Biomedical Engineering. World Congress 2009. Diagnostic and Therapeutic Instrumentation, Clinical Engineering* (pp. 81–84). Presented at the World Congress on Medical Physics and Biomedical Engineering, Munich, Germany: Springer-Verlag. doi:10.1007/978-3-642-03885-3_23
- Nascimento, L. N., & Calil, S. J. (2009b). Allocation of Medical Equipment Costs to Medical Procedures. In J. VanderSloten, P. Verdonck, M. Nyssen, & J. Hauelsen (Eds.), *4th European Conference of the International Federation for Medical and Biological Engineering* (Vol. 22, pp. 1730–1733). New York: Springer.
- Novičević, B. M., & Ljilja, A. (1999). Total quality management and activity-based costing. *Facta Universitatis. Series: Philosophy and Sociology*, 1(7), 1–8.
- Öker, F., & Adigüzel, H. (2010). Time-driven activity-based costing: An implementation in a manufacturing company. *Journal of Corporate Accounting & Finance*, 22(1), 75–92. doi:10.1002/jcaf.20646
- Özbayrak, M., Akgün, M., & Türker, A. K. (2004). Activity-based cost estimation in a push/pull advanced manufacturing system. *International Journal of Production Economics*, 87(1), 49–65. doi:10.1016/S0925-5273(03)00067-7
- Oztaysi, B., Baysan, S., & Dursun, P. (2007). A novel approach for economic-justification of RFID technology in courier sector: a real-life case study. In *1st Annual RFID Eurasia Eurasia Conference, 2007* (pp. 327–331). Presented at the 1st Annual RFID Eurasia Eurasia Conference, Istanbul, Turkey.
- Pernot, E., Roodhooft, F., & Van den Abbeele, A. (2007). Time-Driven Activity-Based Costing for Inter-Library Services: A Case Study in a University. *The Journal of Academic Librarianship*, 33(5), 551–560. doi:10.1016/j.jacalib.2007.06.001
- Raaij, E. M. van. (2005). The strategic value of customer profitability analysis. *Marketing Intelligence & Planning*, 23(4), 372–381. doi:10.1108/02634500510603474
- Ratnatunga, J., Tse, M. S. C., & Balachandran, K. R. (2012). Cost Management in Sri Lanka: A Case Study on Volume, Activity and Time as Cost Drivers. *The International Journal of Accounting*, 47(3), 281–301. doi:10.1016/j.intacc.2012.07.001
- Ratnatunga, J., & Waldmann, E. (2010). Transparent Costing: Has the emperor got clothes? *Accounting Forum*, 34(3–4), 196–210. doi:10.1016/j.accfor.2010.08.005
- Reddy, K., Venter, H., & Olivier, M. (2011). Using time-driven activity-based costing to manage digital forensic readiness in large organisations. *Information Systems Frontiers*, 1–17. doi:10.1007/s10796-011-9333-x
- Ruiz de Arbulo, P., Fortuny, J., García, J., Díaz de Basurto, P., & Zarrabeitia, E. (2012). Innovation in Cost Management. A Comparison Between Time-Driven Activity-Based Costing (TDABC) and Value Stream Costing (VSC) in an Auto-Parts Factory. In S. P. Sethi, M. Bogataj, & L. Ros-McDonnell (Eds.), *Industrial Engineering: Innovative Networks* (pp. 121–128). Springer London. Retrieved from <http://www.springerlink.com/content/r30ppl004366ml97/abstract/>
- Sherratt, M. (2005). Time-Driven Activity-Based Costing. *Harvard Business Review*, 83(2), 144–144.

- Siguenza-Guzman, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2013). Using Time-Driven Activity-Based Costing to Support Library Management Decisions: A Case Study for Lending and Returning Processes. *Library Quarterly*, 84 (1), 76-98.
- Somapa, S., Cools, M., & Dullaert, W. (2010). Time driven activity based costing in a small road transport and logistics company (pp. 281–300). Presented at the Vervoerslogistieke Werkdagen 2010, 18-19 November 2010, Antwerp, Belgium / Witlox, F.J.A.; et al., Antwerp, Belgium: University Press. Retrieved from <http://hdl.handle.net/10067/866250151162165141>
- Somapa, S., Cools, M., & Dullaert, W. (2011). The Development of Time Driven Activity-Based Costing Models: A Case Study in a Road Transport and Logistics Company. In *Current Issues in Shipping, Ports and Logistics* (pp. 431–445). Asp / Vubpress / Upa.
- Stout, D., & Propri, J. (2011). Implementing Time-Driven Activity-Based Costing at a Medium-Sized Electronics Company. *Management Accounting Quarterly*, 12(3), 1–11.
- Stouthuysen, K., Swiggers, M., Reheul, A.-M., & Roodhooft, F. (2010). Time-driven activity-based costing for a library acquisition process: A case study in a Belgian University. *Library Collections, Acquisitions, and Technical Services*, 34(2-3), 83–91. doi:16/j.lcats.2010.05.003
- Sudarsan, P. K. (2006). A resource allocation model for university libraries in India. *Bottom Line: Managing Library Finances*, The, 19(3), 103–110. doi:10.1108/08880450610682527
- Szucs, D., Bauwens, J., Alfonso, J., Holtuis, N., Von Knorring, J., & Demey, J. (2009). Development of a Time-Driven Activity-Based Costing Model for Assessing the Societal Acquisition Cost of Erythrocyte Concentrates in Europe. *Haematologica-the Hematology Journal*, 94, 51–52.
- Szychta, A. (2010). Time-Driven Activity-Based Costing in Service Industries. *Social Sciences*, 67(1), 49–60.
- Tse, M. S. C., & Gong, M. Z. (2009). Recognition of idle resources in time-driven activity-based costing and resource consumption accounting models. *Journal of Applied Management Accounting Research*, 7(2), 41–54.
- Varila, M., Seppänen, M., & Suomala, P. (2007). Detailed cost modelling: a case study in warehouse logistics. *International Journal of Physical Distribution & Logistics Management*, 37(3), 184–200. doi:10.1108/09600030710742416
- Verlinden, B., Duflou, J. R., Collin, P., & Cattrysse, D. (2008). Cost estimation for sheet metal parts using multiple regression and artificial neural networks: A case study. *International Journal of Production Economics*, 111(2), 484–492. doi:10.1016/j.ijpe.2007.02.004
- Wegmann, G. (2010). Compared Activity-Based Costing Case Studies in the Information System Departments of Two Groups in France: A Strategic Management Accounting Approach. In *Proceedings of Business and Information* (Vol. 7, p. 19). Presented at the International Conference on Business and Information. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1620797
- Wegmann, G., & Nozile, S. (2009). The activity-based costing method developments: state-of-the art and case study. *The IUP Journal of Accounting Research and Audit Practices*, 8(1), 7–22. doi:oai:halshs.archives-ouvertes.fr:halshs-00482137
- Wise, K., & Perushek, D. E. (1996). Linear goal programming for academic library acquisitions allocations. *Library Acquisitions: Practice & Theory*, 20(3), 311–327. doi:10.1016/0364-6408(96)00065-8
- Yilmaz, R. (2008a). Creating The Profit Focused Organization Using Time-Driven Activity Based Costing (p. 8). Presented at the Business & Economics Conferences (EABR) and Teaching & Education Conferences (TLC), Salzburg, Austria. Retrieved from <http://rifatyilmaz.net/SLZ08-191.pdf>
- Yilmaz, R. (2008b). The Relations Between Balanced Scorecard and Other Organizational Development Applications. In *Business & Economics Conferences (EABR) and Teaching & Education Conferences (TLC)* (p. 10). Presented at the Business & Economics Conferences (EABR) and Teaching & Education Conferences (TLC), Salzburg, Austria.

Chapter 4: Time-Driven Activity-Based Costing for Lending and Returning Processes

Siguenza-Guzman, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2014). Using Time-Driven Activity-Based Costing to Support Library Management Decisions: A Case Study for Lending and Returning Processes. Library Quarterly: Information, Community, Policy, 84(1), 1–23.

This chapter presents a detailed case study and corresponding analysis conducted using Time-Driven Activity Based Costing (TDABC) on lending and returning processes. The applicability of TDABC in academic libraries is illustrated with special attention to large-scale libraries. This chapter firstly provides a theoretical background of TDABC, including its main characteristics and limitations. Then, the TDABC implementation in the case study is analyzed, identifying key benefits and deployment limitations faced during the process. Finally, some conclusions and recommendations are also provided.

Apart from typographical adjustments, the content of this chapter is identical to the content of the published paper quoted above; where necessary, additional information or remarks are added in footnotes. The layout is adapted for consistency throughout this dissertation. Some redundancy with other chapters is unavoidable as an academic article needs its own introductory sections. This, however, entails the advantage that the chapter can be read separately.

Abstract

With the rapid increase in demand of new digital services, the high cost of information, and the dramatic economic slowdown, libraries have been pressured to improve their services at lower costs. To cope with these conditions, library managers must improve their knowledge and understanding of cost behavior, as well as be aware of the different costs involved in the library. Time-Driven Activity-Based Costing (TDABC) is a cost management technique that allows developing accurate cost information on a wide range of activities. Few case studies have been implemented in libraries regarding very specific processes such as inter-library loan and acquisition processes. More research is still needed to determine whether TDABC is useful and feasible to implement for a more extensive set of library activities. Through an analysis performed at an academic library in Belgium, this document introduces TDABC as a useful method for supporting lending and returning processes.

Contributions of the first author

The first author's contributions are: the literature study on TDABC and its implementation in libraries. It also includes data collection, TDABC implementation, implications, and conclusions.

4.1 Introduction

Due to the recent economic crisis, the high cost of information, and the rising demand of services and information resources, libraries have been required to shift budgeting and spending priorities. As a consequence, several decisions have been made, such as cutting collection budgets, eliminating budgets for travel or conferences, freezing salaries, finding new ways to fund programs, and moving from physical to digital collections (McKendrick, 2011; Sudarsan, 2006). This evolution has forced libraries to prioritize their spending and minimize their costs, concentrating on key success factors such as cost efficiency, quality and innovations (Novičević & Ljilja, 1999; Blixrud, 2003, p. 15).

Library managers in these difficult circumstances are required to increase their understanding of library activities and their related costs in order to justify resource requirements and the creation of new services or face budget reductions. To do so, they must rely on valid information about the library processes and cost estimations, as well as differentiate the kind of “products” or “services” libraries provide to customers. For instance, there are no tangible products in libraries (except for scanning and photocopying) and the primary products are a wide range of services. Several studies on cost analysis and resource allocation for library services have been developed over the past forty years (Kaplan & Cooper, 1998; Rouse, 1975), in which traditional costing systems have been mainly used (Kaplan & Cooper, 1998).

This article introduces Time-Driven Activity-Based Costing (TDABC) as a useful costing system for librarians and library managers who want to perform a cost analysis in a simple and accurate manner. TDABC, which was initially developed for manufacturing processes to overcome the difficulties presented by traditional costing systems, is gaining special attention in academic libraries. This is because TDABC is a fast, accurate, and easy-to-understand method that only requires two parameters: the unit cost of supplying resource capacity and an estimated time required to perform an activity (Kaplan & Anderson, 2007b). By implementing TDABC in libraries, key benefits are provided, such as the possibility of disaggregating values per activity to identify non-value added activities; benchmarking different scenarios to adapt best practices for performance improvement; and justifying decisions and choices such as staff recruitment, training and new service development (Siguenza-Guzman et al., 2013). This investigation presents a case study of the Loan and Return processes at the Arenberg Library of the Katholieke Universiteit Leuven (KU Leuven) to illustrate the applicability of TDABC in academic libraries with special attention to large-scale libraries. The remainder of this article is organized as follows. The next section presents the theoretical background of TDABC and its main characteristics and limitations. Then, the implementation of TDABC in a case study is analyzed, identifying key benefits and deployment limitations faced during the process. Finally, some conclusions and recommendations for future work are given in the last section.

4.2 Theoretical background: TDABC

The most well-known costing system is the so-called traditional costing, which consists of a single and static cost rate for allocating indirect costs of different processes to cost objects such as products or services (Kaplan & Cooper, 1998). It works well when used in specific scenarios, such as in stable environments with small or fixed indirect costs, but it leads to inaccurate total product cost estimations in more complex environments (Kaplan & Cooper, 1998; Tse & Gong, 2009). As a result of the current wide offering of library products and services, these inaccuracies become critical in the ability to accurately describe the complexity of the cost structure (Tse & Gong, 2009).

Activity-Based Costing (ABC) is an alternative costing technique specially promoted by Robert S. Kaplan and Robin Cooper (1988) in the mid-1980s. In the case of libraries, ABC performs a more accurate and efficient treatment of activity costs compared to traditional costing due to its accuracy in allocating indirect costs to different activities (Ellis-Newman, Izan, & Robinson, 1996; Ellis-Newman & Robinson, 1998; Goddard & Ooi, 1998; Ellis-Newman, 2003; Ching, Leung, Fidow, &

Huang, 2008; Novak, Paulos, & Clair, 2011). However, in practice, ABC is time consuming and costly, which is mainly a result of data collection performed by means of interviews (Kaplan & Anderson, 2004). As a consequence, several companies have ceased updating their systems, and in some cases they have substituted more efficient approaches such as TDABC (Yilmaz, 2008, p. 8; Wegmann & Nozile, 2009).

TDABC is a new ABC approach developed by Robert S. Kaplan and Steven R. Anderson (2004) to overcome the difficulties of implementing and updating ABC systems. TDABC assigns resource costs directly to the cost objects using two easy-to-obtain sets of estimates: (1) the cost per time unit of supplying resource capacity to the activities and (2) an estimate of the time units required to perform an activity (Kaplan & Anderson, 2004). To calculate the cost of activities under a TDABC system, six steps typically need to be performed (Everaert, Bruggeman, Sarens, Anderson, & Levant, 2008). These steps are illustrated in Table 4.1.

Table 4.1: Time-Driven Activity-Based Costing steps (Everaert et al., 2008)

Step	Description
1	Identify the services or activities
2	Estimate the total cost of each resource group
3	Estimate the practical time capacity of each resource group
4	Calculate the unit cost of each resource group
5	Determine the estimated time for each activity
6	Multiply the unit cost of each resource group by the estimated time for the activity

TDABC starts estimating the cost of supplying capacity by identifying the different services or activities, their cost and their practical capacity. Then, the unit cost of each resource group is gathered by dividing the total cost by the practical capacity. The total cost of supplying resource capacity is defined as the cost of all the resources supplied to this department or process (e.g., staff, supervision, occupancy, equipment, technology, and infrastructure). Practical capacity is defined as the amount of time that employees work without idle time (Kaplan & Anderson, 2007a). There are two ways to estimate practical time capacity: (1) assuming an 80 percent of theoretical capacity for people (excluding breaks, arrival and departure, communication, training, meetings, chitchat, etc.), and an 85 percent capacity for machines (excluding maintenance, repair, and scheduling fluctuations); and (2) calculating the real values adjusted for the institution (e.g., available working hours, excluding holidays, meeting and training hours; (Kaplan & Anderson, 2007b)).

Once the cost of supplying capacity has been calculated, the estimated time for each activity is determined. This value can be obtained through interviews with employees or by direct observation; no additional surveys are required. Authors argue that precision is not critical, that a rough accuracy is sufficient because gross inaccuracies will be revealed either in unexpected surpluses or shortages of committed resources (Kaplan & Anderson, 2007b). Unlike ABC, this value refers to the time that an employee spends doing an activity and not the percentage of time that it takes to complete one unit of that activity. In addition, through a simple time equation, it is possible to represent all possible combinations of activities (e.g., different types of products do not necessarily require the same amount of time to be produced). For each activity, costing equations are calculated based on the time required to perform an activity (Yilmaz, 2008). This time is computed by time equations, which are the sum of individual activity times. Using these equations, it is possible to combine all the activities involved into one process with only one time equation. They are represented with the following expression (Kaplan & Anderson, 2007b):

$$\text{Time required to perform an activity} = (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \dots + \beta_i X_i)$$

Where β_0 is the standard time to perform the basic activity (e.g., 5 minutes), β_1 is the estimated time for the incremental activity i (e.g. time required for the hand-in-robot to allocate the item in

the correct box = 0.5 minutes), and X_i is the quantity of incremental activity i (e.g., number of items per loan = 2). Finally, costs are assigned to the cost object by multiplying the cost per time unit of the resources by the estimate of the time required for performing the activities.

4.3 TDABC in libraries

Little is known about applying TDABC in libraries. The first TDABC approach, given by Eli Pernot, Filip Roodhooft, and Alexandra Van den Abbeele (2007), uses TDABC for calculating interlibrary loan (ILL) costs. The paper offers a useful technique to reduce ILL resource costs or renegotiate ILL service prices. The authors conclude that TDABC could reduce the cost management of all library services because librarians can take appropriate actions to decrease the time needed for specific customer's requests. On the other hand, Kristof Stouthuysen and colleagues (2010) focus their analysis on the acquisition process. They describe TDABC as a useful system for small- to medium-size academic libraries. The authors define TDABC as a costing system that assists library managers in visualizing the acquisition process efficiencies and capacity utilization, leading to potential cost efficiencies. They also state that the study can be applied to complex or digitalized acquisition environments. These initial investigations show promising possibilities of using TDABC to provide accurate information on library activities. However, these studies have been applied to very specific settings, studying only particular processes with cases in small- and medium-size academic libraries. More research is still needed to identify whether the technique of TDABC is useful and feasible to implement for a more extensive set of library activities. In this context, the main objective of this article is to study whether TDABC can support loan and return processes in a large academic library.

4.4 Case Study of Loan and Return Processes

A case study using TDABC was performed in the Circulation Department at the Arenberg Campus Library (CBA - Campusbibliotheek Arenberg) of the KU Leuven in Belgium. We focus on this unit because it is considered one of the most important departments of the library. Although nowadays digital libraries are becoming stronger, physical processes such as loan and return processes are still considered crucial activities within libraries because, as McKendrick (2011, p. 4) states: *"print still commands a lion's share of annual budgets."* The Circulation Department involves all the services related to lending processes, such as loaning, returning, reserving, renewing, paying fines, and providing basic reference services.

CBA is operated by approximately 20.5 full-time-equivalent employees (FTE), who give support to about 10,000 potential customers (Dekeyser & Holans, 2003). To improve its cost efficiency, CBA has been obliged to use new technologies and to automate repetitive processes. In the case of the Circulation Department, the library has implemented lending and returning robots (lending robots named "chicos," as in "check in, check out"). With each chico robot, customers can borrow or return the items without any assistance. Alternatively, customers can also return the materials with a "hand-in-robot" that allows the return of items during hours the library is open without entering the library. The technology used for interacting with the robots and tagging the items is a Radio Frequency IDentification (RFID) system. Because of this interesting automation to improve cost efficiency, the processes of lending and returning have been selected to understand whether the decisions made at that time were the most appropriate. In the next sections, we illustrate the implementation of TDABC in the loaning and returning processes at CBA by applying six steps identified by Patricia Everaert and colleagues (2008).

4.4.1 Step 1: Identifying the services or activities

The first step is to identify the main activities of the Circulation Department and the role that each staff member has in these activities. In order to do this, a round of interviews with the head of the library and the main desk staff (i.e., people in charge of those processes) was conducted.

Three main activities in the Circulation Department, WBIB lending (Wetenschappelijke Bibliotheek – scientific library), WMAG lending (Wetenschappelijke MAGazijn – scientific warehouse), and returning, were identified. Lending is the process of allowing users to borrow one or more items temporarily from the library. In the case of the CBA, the lending processes can be triggered with two types of materials: (1) WBIB items, which are bibliographic material (e.g., books or journals) located in open shelves, are directly available for customers without any assistance from the librarian. (2) WMAG items are only accessible to the library staff, because they are stored in compact shelves located in the basement. When a customer needs items from this repository, an online request must be filled. Returning is the process wherein customers return borrowed items to the library.

A second round of interviews was performed to obtain specific details about the different sub-activities of each process. This additional information was used to build flow charts of the activity sequences in the processes. It is important to remark that the least accurate flow occurred when superiors provided descriptions or when librarians presented a printed report of their estimations, which supports the findings of Eddy Cardinaels and Eva Labro (2008).

MS Visio and MS Excel were used to create a graphical representation of the tabular information. In each of the following figures, the beginning of each process is represented by a closed circle ●. Each figure is divided by horizontal lines to represent each of the actors involved in the process (e.g., customer, closed-stack responsible, main desk). These lines allow one to easily identify the different actors and resources involved in every specific activity. A diamond ◆ depicts the different options that a process has in a specific moment, such as the two possibilities to bring an item from the closed stack if it is an evening or a day shift. The rounded rectangles represent the different activities, the average time consumed as well as its incurred cost. The end of a process is represented by the symbol ○.

In the case of WBIB items (Figure 4.1), the customer and the main desk are the two actors involved. The process starts when a customer consults the catalog to get an item. It is possible to find the physical place of the item by using the locator (i.e., a web system helps to locate library items); otherwise, the customer can go directly to the corresponding shelf. If the customer decides to borrow the item(s), the customer then puts them on the chico robot. If the customer has outstanding fines, the transaction is not performed until the fine is paid through an electronic transaction or in cash at the main desk. Finally, the customer can also print a receipt, which includes the details of the borrowed items.

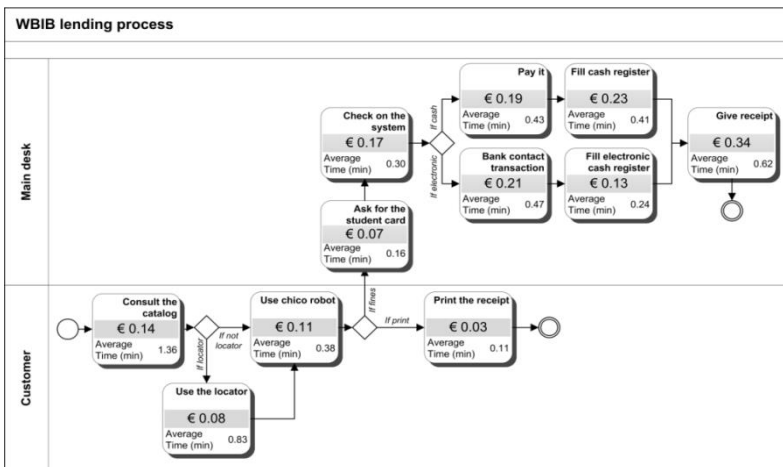


Figure 4.1: WBIB Lending Process

In the WMAG Lending Process (Figure 4.2), four actors are involved: the customer, the main desk, a student library employee (SLE, i.e., a student hired to perform reshelving and classification activities), and the closed stack responsible. This process starts similarly to the previous one. However, in this case, the locator does not appear because it is not an open-shelf item. The customer requests the item online and receives it from the main desk. CBA has two kinds of staff shifts. In the mornings, a librarian works in the closed stack section and is responsible for sending the item to the main desk. During the evenings, the closed stack area is closed. Hence, an SLE goes to the closed stack and brings the item to the main desk. In any case, employees will ensure that the item is tagged (RFID system) before giving it to the customer. At this moment, the customer has two alternatives: (1) he or she can borrow the item, which results in the same procedure as the WBIB Lending, or (2) he or she can return the item, in which case the librarian deletes the request. If the item has also been requested by another customer, the librarian sends an e-mail message to the other customer, prints the request, and puts the item on a special shelf. Otherwise, if there is not a request for the item, the librarian returns the item to the closed stack in order to be reshelved.

The Returning Process (Figure 4.3) is triggered by the customer returning the item. In this process, two actors are involved: the customer and an SLE. The customer has two possibilities of returning: (1) He or she can return the item by the chico robot and print the receipt as proof of the returning. The robot has a plate where the customer puts on the borrowed items in order to make them recognizable for the computer. The customer can place a maximum of five items on the plate at one time. (2) He or she can return the item by hand in the robot and print the receipt. Using this machine, the customer is required to put the items one by one.

In the first case, during the evenings an SLE goes to the hand-in-robot with the book truck and returns the items. The objective of this activity is to accelerate the items' classification by the corresponding cluster (i.e., book collection divisions). In the second case, this activity is not necessary because the items are already sorted when the customer puts the items in the hand-in-robot. In any case, once the items have been classified by the hand-in-robot, an SLE will sort the items in the cluster; and then reshelv the items.

4.4.1 Step 2: Estimating the total cost of resource groups

The total cost for the processes of lending and returning consists of four main direct costs: staff costs, machines costs, library management systems (LMS) costs and SLE costs. Due to differences in the staff salaries performing the activities, the average salary is used for the staff cost. Each salary is calculated by the monthly gross salary plus the social security contribution. This represents the full cost of an employee. According to the head of the library, the total number of personnel assigned to the three processes represents 1.5 FTE. This value corresponds to several people working different percentages of their time as a comparable number to full-time employees. It corresponds to about € 4,110 on a monthly basis.

The cost of an SLE is about € 441.37 monthly. The head of the library reported having an average of five students per month, each working forty hours (total job students = $5 \times 40 \frac{h}{m} = 200 \frac{h}{m}$). If we want to report the equivalence in FTE, we should consider that a staff member works thirty-eight hours/week (equals one FTE). If we assume four weeks per month, the corresponding FTE for SLE is $1.32 \left(FTE = 200 \frac{h}{m} \div \left(38 \frac{h}{w} \times 4 \frac{w}{m} \right) = 1.32 FTE \right)$, equivalent to € 582.60 monthly.

Loans and returns are mainly done by machines and the yearly cost associated with their maintenance, repair and inspection is about € 30,400 (including VAT [value added tax]). The value includes the costs associated with the maintenance of the RFID robots (chico and hand-in-robot) and the gate antennas (security). The yearly cost incurred with LMS is about € 17,050.

Based on the library accounting reports, there are two main types of indirect costs: (1) € 3,000 for machine overhead costs on a yearly basis (e.g., computer supplies, hardware, software) and (2) € 191,060 for staff overhead costs on a yearly basis (e.g., management, secretary, accounting, training

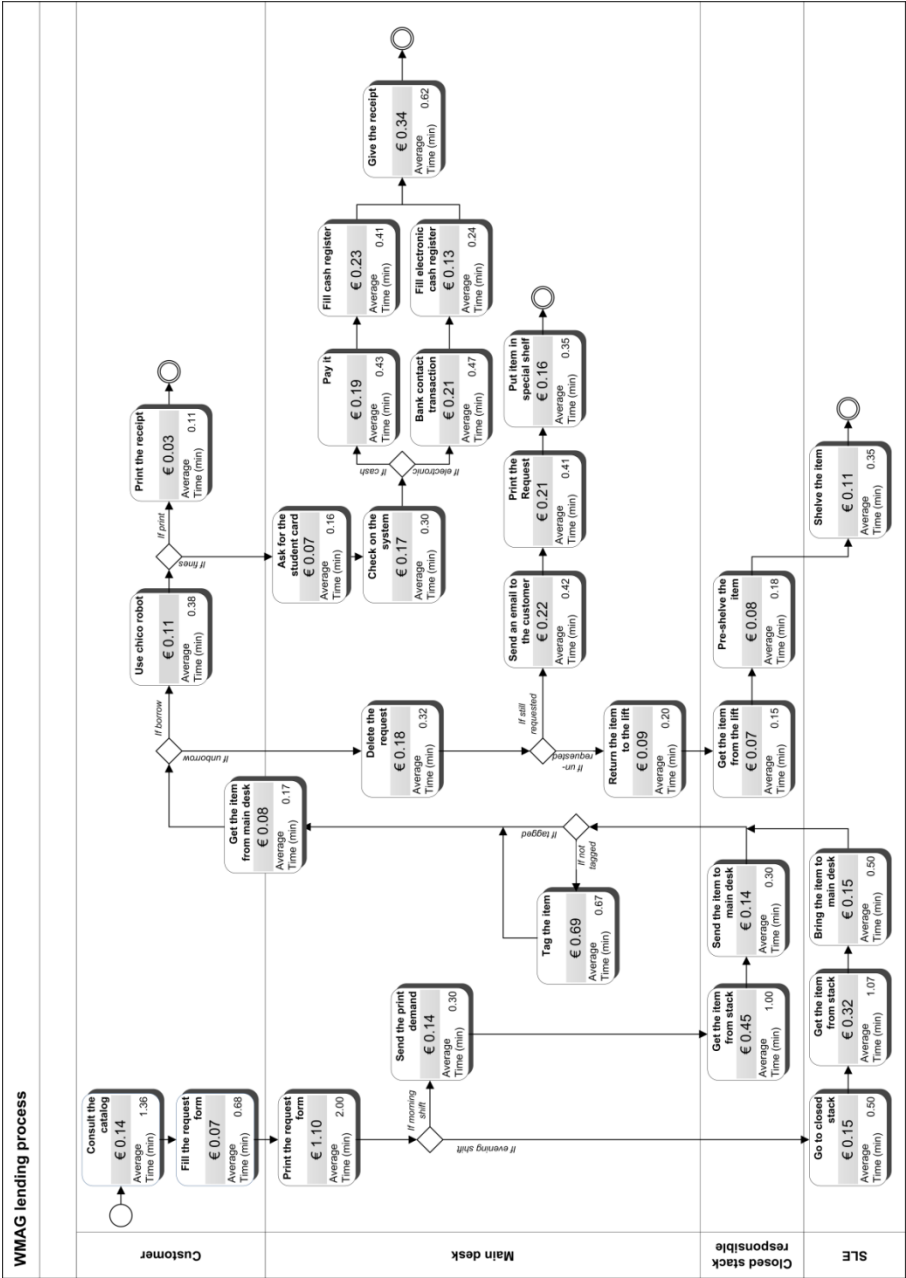


Figure 4.2: WMAG Lending Process

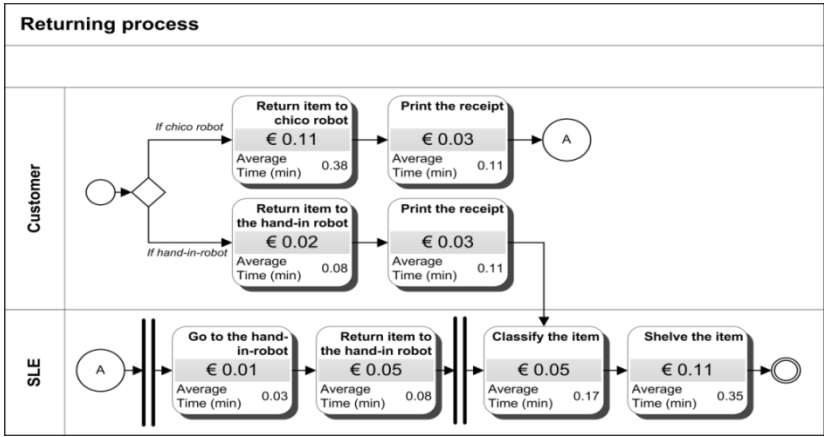


Figure 4.3: Returning Process

meetings staff, stationery material).⁵ Because staff and SLE are associated with the second overhead, this cost is divided by the total number of FTE (FTE staff + FTE job student = 20.5 + 1.32 = 21.82), resulting in a yearly overhead of € 8,756 per FTE (€ 730 per month).

4.4.2 Step 3: Estimating the practical time capacity of each resource group

There are two ways to get the value of the practical time capacity of each practical full capacity is about 80 percent for people and 85 percent for machines; and (2) calculating or counting the real values according to the library situation (Kaplan & Anderson, 2007b). In order to simplify the study, the first option has been selected.

In the case of the machines and LMS, the practical capacity is set equal to the time that library is open, that is, for weekdays from 9 a.m. until 10 p.m. and for Saturdays from 9 a.m. until 1 p.m. (Dekeyser & Holans, 2003). This means that on a theoretical basis machines and LMS are available during sixty-nine hours per week. Assuming fifty-two weeks per year, the practical capacity for machines and LMS is 182,988 minutes/year $\left(85\% \times 69 \frac{\text{hours}}{\text{week}} \times 60 \frac{\text{min}}{\text{hour}} \times 52 \frac{\text{weeks}}{\text{year}}\right)$.

For staff capacity, thirty-eight hours per week are accounted for the theoretical capacity. This results in 7,296 minutes/month $\left(80\% \times 38 \frac{\text{hours}}{\text{week}} \times 4 \frac{\text{weeks}}{\text{month}} \times 60 \frac{\text{min}}{\text{hour}}\right)$. Considering 1.5 FTE for the lending and returning processes, the practical capacity for staff is 10,944 min/month.

Finally, regarding the SLE capacity, the theoretical capacity is forty hours/month (according to regulations by the law); the practical capacity for each SLE is 1,920 minutes/month $\left(80\% \times 40 \frac{\text{hours}}{\text{month}} \times 60 \frac{\text{min}}{\text{hour}}\right)$. Considering 1.32 FTE for SLE, the practical capacity is 2,534.4 minutes/month.

⁵ There are certain values that are not required to be included as indirect costs because they are paid by the university and not charged to the library (e.g., electricity, heating, transportation, telephone; Ellis-Newman & Robinson, 1998). Machines depreciation costs are not included since this equipment has been in use for several years already.

4.4.3 Step 4: Calculating the unit cost of each resource group

The cost per unit time is calculated by dividing the total cost of the resource (step 2) by the practical capacity (step 3). The machine overhead is added to the machines costs and LMS, and the staff overhead is added to the staff and the SLE costs. The resulting costs are presented in Table 4.2.

Table 4.2: Unit cost per resource group

Cost Type	Calculations	Cost per Minute (€/min)
Machines	$= (\text{€ } 30,400/182,988) + (\text{€ } 3,000/182,988) = .17 + .01$.18
Library management systems	$= (\text{€ } 17,045/182,988) + (\text{€ } 3,000/182,988) = .09 + .01$.10
Staff labor	$= (\text{€ } 4,110/10,944) + (\text{€ } 730/10,944) = 0.38 + 0.07$.45
Student library employee	$= (\text{€ } 582,60/2,534.4) + (\text{€ } 730/10,944) = 0.23 + 0.07$.30

4.4.4 Step 5: Determining the time estimation for each activity

The required time to perform an activity was gathered through direct observation. The data collection was conducted multiple times using a stopwatch during several days at different hours in the first semester of 2010, and the results were validated through a second data collection in the following semester in order to avoid possible biases. Since we made a large number of observations, the average time of each activity was taken as reference to facilitate the calculations.

4.4.4.1 WBIB Lending Process

A customer consults the catalog to get an item (1.36 minutes), and then the physical place of the item can be identified by using the locator (0.83 minute). If the customer decides to borrow the item(s), the customer puts it (them) on the chico robot (0.38 minute). In case the customer has outstanding fines, the transaction cannot be performed until the fine is paid by an electronic transaction (0.71 minute) or in cash (0.84 minute) at the main desk. In any case, the librarian will ask for the student card to check on the system the value to be paid (0.46 minute), and will give a receipt to the customer as proof of payment (0.62 minute). Finally, if the customer desires, he or she can also print the borrow register (0.11 minute). The resulting equation is as follows:

$$\begin{aligned}
 &\text{WBIB Lending Process} \\
 &= 1.36 \times (\text{number of items}) + 0.83 \times (\text{number of items}) \{if \text{ locator}\} \\
 &+ 0.38 + [(1.08 + 0.71\{if \text{ electronic}\} + 0.84\{if \text{ cash}\})\{if \text{ fines}\}] \\
 &+ 0.11\{if \text{ print receipt}\}
 \end{aligned} \quad (1)$$

There are certain parameters that can influence the formula (e.g. if the customer uses the locator or goes directly to the shelf). These situations are represented by dummy variables (Boolean values; Everaert & Bruggeman, 2007). The dummy variable is zero when the optional activity is not used in a specific situation. Otherwise, the dummy is one when the activity is part of a particular process. In equation (1), dummy variables are $\{if \text{ locator}\}$, $\{if \text{ electronic}\}$, $\{if \text{ cash}\}$, $\{if \text{ fines}\}$, and $\{if \text{ print receipt}\}$.

4.4.4.2 WMAG Lending Process

The WMAG Lending Process starts similarly to that shown in equation (1) (1.36 minutes). Subsequently, the customer requests the item (0.68 minute) which is automatically printed by the main desk (2.00 minutes). In the morning shift, the printed request is sent to the closed stack by the lift (0.30 minute), the closed stack responsible gets the item from the stack (1.00 minute) and sends it back to main desk (0.30 minute). In the evening shift, an SLE goes from the main desk to the closed stack (0.50 minute), gets the item (1.07 minutes) and carries it to the main desk (0.50 minute). If the item is not tagged, the employee will tag it (0.67 minute) and then give it to the customer (0.17 minute). An individual tag costs € 0.30, including VAT.

Then the customer has two alternatives: (1) *Borrow the item*: put the item on the chico robot and perform the borrowing procedure (0.38 minute). The transaction will not be made unless all outstanding fines are paid, similarly to equation (1). The process finishes by printing the receipt as a proof to the lending (0.11 minute). (2) *Return the item*: the librarian deletes the request (0.32 minute). If the item has another request, the librarian sends an email message to the customer (0.42 minute), prints the request (0.41 minute), and puts the item in a special shelf (0.35 minute). Otherwise, the librarian returns the item to the closed stack (0.35 minute) for it to be reshelfed (0.53 minute).

$$\begin{aligned}
 \text{WMAG Lending Process} = & 1.36 \times (\text{number of items}) + 0.68 \times (\text{number of items}) + \\
 & 2.00 + (0.30 + 1.00 \times (\text{number of items}) + 0.30)\{\text{if morning shift}\} + (0.50 + 1.07 \times \\
 & (\text{number of items}) + 0.50)\{\text{if evening shift}\} + 0.67 \times (\text{number of items})\{\text{if not tagged}\} + \quad (2) \\
 & 0.17 + \\
 & (0.38 + [(1.08 + 0.84 \{\text{if electronic}\} + 0.71 \{\text{if cash}\})\{\text{if fines}\}] + \\
 & 0.11\{\text{if print receipt}\})\{\text{if borrow}\} + [0.32 + (0.42 + 0.41 + 0.35)\{\text{if still requested}\} + \\
 & (0.20 + 0.15 + 0.18 + 0.35)\{\text{if not new request}\}] \times (\text{number of items})\{\text{if unborrow}\}
 \end{aligned}$$

4.4.4.3 Returning Process

There are two options to return item(s): (1) by the chico robot (0.38 minute), printing the receipt as a proof to the return (0.11 minute), and leaving the item in the book truck, or (2) by the hand-in-robot (0.08 minute) and printing the receipt (0.11 minute). In the first case, during the evenings, an SLE goes to the hand-in-robot with the book truck (0.03 minute) and returns every item (0.08 minute). In both cases, the hand-in-robot will classify the items by the corresponding cluster, and an SLE will sort the items in the cluster (0.17 minute) and then reshelf the items (0.35 minute). The last two values are calculated using an average batch of items that an SLE classifies per cluster.

$$\begin{aligned}
 \text{Returning Process} = & 0.38\{\text{if chico robot}\} + 0.08(\text{number of items})\{\text{if hand in robot}\} + \\
 & 0.11\{\text{if print receipt}\} + [0.03 + 0.08 \times (\text{number of items})]\{\text{if chico's robot}\} + \quad (3) \\
 & (0.17 + 0.35) \times (\text{number of items})
 \end{aligned}$$

4.4.5 Step 6: Multiplying the unit cost of each resource group by the time estimate for the activity

Eventually, a cost table is built by multiplying the unit cost per time and the time needed for the activity. Tables 4.4-4.6 present the costs incurred in each activity. The first column in each of these tables lists the activities identified in the process, and the second column shows the average time for each event. The third column indicates the accumulated costs of each resource group, and the fourth column calculates the resulting cost incurred in the activity. The fifth column describes the dummy variable conditioning the activity, and the sixth column includes the resource group

involved in each activity. Each table is divided by standard and optional activities to separate the values influenced by the dummy variables.

To calculate the total cost of a sub-process, one first identifies the different activities that appear in this specific situation. The fifth column of the costing tables that contains the dummy variables helps us to identify which optional activities are going to be used for the calculation. In the bottom part of Table 4.3, three different examples of sub-processes are included to illustrate how the total cost is calculated. Case A represents the most common situation of a customer taking the item from the shelf, borrowing it, and finally printing the receipt. Case B and C correspond to a customer who is trying to borrow an item, but first he has to pay a fine of previous transactions. In case B, the customer pays the fine through an electronic transaction, whereas in case C he pays the fine in cash.

Based on these costing tables a costing analysis process, by means of a what-if analysis, can be performed. For instance, it is possible to analyze the cost of returning three items through all possible cases: (1) by the chico robot, (2) by the hand-in-robot, and (3) manually with the assistance of a librarian.

If a customer returns the items through the chico robot, this value (0.38 minute) is not multiplied by the number of items because these machines allow the customer to return up to five items at the same time on the plate. Then a receipt could be printed as proof of returning (0.11 minute). Those values are multiplied by the cost that represents the maintenance of the machines (€ 0.18) and LMS (€ 0.10). Table 4.6 provides an overview of this example.

If the items are returned by the hand-in-robot, the customer is required to place the items one by one onto the machine. In this case, that time value (0.08 minute) is multiplied by the number of items the customer returns. The final receipt will be only one, even if the customer returns many more items. The results can be seen in Table 4.7.

If the items are returned manually with a librarian, the customer gives the items to the main desk. The librarian scans each item by hand to enter into the system (0.20 minute) and then rewrites the RFID tag to specify that the item is in the library again (0.32 minute). In order to do this the librarian places the item in the RFID station. Finally, the employee leaves the items in the book truck. Table 4.8 contains the results.

4.5 Implications

The TDABC analysis provides many insights into the costs of the lending and returning processes. This, in turn, leads to several implications and recommendations. From the example on the returning process, it is evident that the time and cost of returning three items manually is very high in comparison with the same activity performed by the robots. If the items are returned by the chico robot, we obtain a reduction in costs of 47 percent. If this task is performed through the hand-in-robot, the cost is 20 percent less in comparison to the chico robot. Based on these figures it is recommended to automate as much as possible the returning process, as this will lead to significant cost reductions.

Although, most of the processes in the Circulation Department have been automated, there are still some activities that can be improved. For instance, in table 3 the total cost of a WBIB lending process is analyzed through three different cases. The TDABC analysis shows that paying fines electronically is slightly less expensive than paying in cash (6 percent), as in both cases the assistance of a librarian is required. However, if a customer does not have to pay a fine, the cost is reduced by 76 percent. Based on these findings, potential improvements can be undertaken, such as performing awareness-raising campaigns about returning items on time or paying fines at the time of enrollment at the University.

Table 4.3: WBIB Lending Process Cost Table

Activity	Average Time (Minutes)	Cost (€/Minute)	Cost (€)	Dummy Variable	Resources	Activity Symbol
<i>Standard activities:</i>						
Consult the catalog	1.36	.10	.14		LMS	a
Use chico robot	.38	.28	.11		Chico robot + LMS	b
Subtotal	1.74		.25			
<i>Optional activities:</i>						
Use the locator	.83	.10	.08	If locator	LMS	c
Print the receipt	.11	.28	.03	If print receipt	Chico robot + LMS	d
Ask for the student card	.16	.45	.07	If fines	Main desk	e
Check on the system	.30	.55	.17	If fines	Main desk + LMS	f
Pay it	.43	.45	.19	If fines, if cash	Main desk	g
Fill cash register	.41	.55	.23	If fines, if cash	Main desk + LMS	h
Bank contact transaction	.47	.45	.21	If fines, if electronic	Main desk	i
Fill electronic cash register	.24	.55	.13	If fines, if electronic	Main desk + LMS	j
Give receipt	.62	.55	.34	If fines	Main desk + LMS	k
Subtotal			
Total			

(A)	$a + b + d \text{ (if print receipt)}$ $1.36 + .38 + .11 \text{ (if print receipt)} = 1.36 + .38 + (.11 \times 1) = 1.85 \text{ minutes}$ $.14 + .11 + .03 \text{ (if print receipt)} = .14 + .11 + (.03 \times 1) = \text{€}.28$
(B)	$a + b + \{[(e + f + k) + [(i + j) \text{ (if electronic)}]] \text{ (if fines)}\}$ $1.36 + .38 + \{[(1.08 + .71 \text{ (if electronic)}) \text{ (if fines)}]\} = 1.36 + .38 + 1.08 + .71 = 3.53 \text{ minutes}$ $.14 + .11 + \{[(.58 + .34 \text{ (if electronic)}) \text{ (if fines)}]\} = .14 + .11 + .58 + .34 = \text{€}1.17$
(C)	$a + b + \{[(e + f + k) + [(g + h) \text{ (if cash)}]] \text{ (if fines)}\}$ $1.36 + .38 + \{[(1.08 + .84 \text{ (if cash)}) \text{ (if fines)}]\} = 1.36 + .38 + 1.08 + .84 = 3.66 \text{ minutes}$ $.14 + .11 + \{[(.58 + .42 \text{ (if cash)}) \text{ (if fines)}]\} = .14 + .11 + .58 + .42 = \text{€}1.25$

Note. – WBIB = (Wetenschappelijke Bibliotheek [open-shelf collection]). LMS = library management systems.

Table 4.4: WMAG Lending Process Cost Table

Activity	Average Time (Minutes)	Cost (€/Minute)	Cost (€)	Dummy Variable	Resources
<i>Standard activities:</i>					
Consult the catalog	1.36	.10	.14		LMS
Fill the request form	.68	.10	.07		LMS
Print the request form	2.00	.55	1.10		Main desk + LMS
Get the item from main desk	.17	.45	.08		Main desk
Use chico robot	.38	.28	.11		Chico robot + LMS
Subtotal	4.59		1.49		
<i>Optional activities:</i>					
Get the item from stack	1.60	.45	.72	If morning shift	Closed stack responsible
Get the item from stack	2.07	.30	.62	if evening shift	SLE
Tag the item	.67	1.03*	.69	If not tagged	Main desk + LMS + machines + tags*
Print the receipt	.11	.28	.03	If borrow	Chico robot + LMS
Ask for the student card	.16	.45	.07	If fines	Main desk
Check on the system	.30	.55	.17	If fines	Main desk + LMS
Pay it	.43	.45	.19	If fines, if cash	Main desk
Fill cash register	.41	.55	.23	If fines, if cash	Main desk + LMS
Bank contact transaction	.47	.45	.21	If fines, if electronic	Main desk
Fill electronic cash register	.24	.55	.13	If fines, if electronic	Main desk + LMS
Give receipt	.62	.55	.34	If fines	Main desk + LMS
Delete the Request	.32	.55	.18	If unborrow	Main desk + LMS
Return the item to the closed stack	.53	.45	.24	If unborrow, if not new request	Closed stack responsible
Shelve the item	.35	.30	.11	If unborrow, if not new request	SLE
Send an email to the customer /					
Print the request	.83	.55	.43	If unborrow, if still requested	Main desk + LMS
Put item in a special shelf	.35	.45	.16	If unborrow, if still requested	Main desk + LMS
Subtotal		
Total		

Note. – WMAG = (Wetenschappelijke MAGazijn [closed stack collection]). LMS = library management systems. SLE = student library employee.

* €1.03 = main desk + LMS + machines + tags = .45 + .10 + .18 + .30 (individual tag costs).

Table 4.5: Returning Process Cost Table

Activity	Average Time (Minutes)	Cost (€/Minute)	Cost (€)	Dummy Variable	Resources
<i>Standard activities:</i>					
Shelve the item	.35	.30	.11		SLE
Subtotal	.35		.11		
<i>Optional activities:</i>					
Return item to chico robot	.38	.28	.11	If chico robot	Chico robot + LMS
Print the receipt	.11	.28	.03	If chico robot or If hand-in-robot	Machines + LMS
Return item to the hand-in-robot	.08	.28	.02	If hand-in-robot	Machines + LMS
Go to the hand-in-robot	.03	.30	.01	If book truck	SLE
Return item to the hand-in-robot	.08	.58	.05	If book truck	SLE + machines + LMS
Classify the item	.17	.30	.05	If book truck	SLE
Subtotal		
Total		

Note. – LMS = library management systems. SLE = student library employee.

Table 4.6: Returning Process through the Chico Robot (Three Items Returned)

Activity	Average Time (min)	Cost (€/min)	Cost (€)	Resources
Shelve the item	.35 x 3	.30	.32	SLE
Return item to chico robot	.38	.28	.11	Chico robot + LMS
Print the receipt	.11	.28	.03	Machines + LMS
Go to the hand-in-robot	.03 x 3	.30	.03	SLE
Return item to the hand-in-robot	.08 x 3	.58	.14	SLE + machines + LMS
Classify the item	.17 x 3	.30	.15	SLE
Total	2.38		.78	

Table 4.7: Returning Process through the Hand-in-Robot (Three Items Returned)

Activity	Average Time (min)	Cost (€/min)	Cost (€)	Resources
Shelve the item	.35 x 3	.30	.32	SLE
Print the receipt	.11	.28	.03	Machines + LMS
Return item to the hand-in-robot	.08 x 3	.55	.13	Machines + LMS
Classify the item	.17 x 3	.30	.15	SLE
TOTAL	1.91		.63	

Table 4.8: Returning Process through the librarian staff

Activity	Average Time (min)	Cost (€/min)	Cost (€)	Resources
Return item manually into the system	0.20*3	0.55	0.33	Main desk + LMS
Rewrite the RFID tag	0.32*3	0.73	0.70	Main desk + Machines + LMS
Print the receipt	0.11	0.55	0.06	Main desk + LMS
Leave the items in the book truck	0.10	0.45	0.05	Main desk
Go to the hand-in-robot	0.03*3	0.30	0.03	SLE
Return item to the hand-in-robot	0.08*3	0.58	0.14	SLE + Machines + LMS
Classify the item	0.17*3	0.30	0.15	SLE
TOTAL	2.61		1.46	

Other what-if analysis can be done between the activities performed by an SLE or a librarian, such as reshelving the materials. An interesting discussion is also to analyze the most expensive activities, for instance, printing requests in the WMAG process (€ 1.10). Taking into account the incurred costs for such an activity and the environmental impact caused, one may consider automating this activity by setting alerts to the system when a new request is performed.

4.6 Conclusions

Because of the current economic conditions of our times and because of limited resources, academic libraries are called to search for efficient methods to balance their limited budgets with

the services provided. Hence, the costs and time consumed by activities, processes and resources are extremely important and of high interest to library managers in identifying non-value-added activities, finding and adapting best practices, and justifying decisions and choices. In this article, a case study of the Time-Driven Activity-Based Costing implementation on the loan and return processes at the Arenberg Campus Library of the KU Leuven was conducted. This case study has illustrated through six simple steps how this method can be used in carrying out a cost analysis in a simple, easy-to-understand and accurate manner.

Several important insights have emerged from the case study. The first important insight is that the amount of time required collecting the duration of activities and to document the activity flows is relatively limited compared to the insights gained from the analysis. The duration of activities was gathered by direct observation since the most accurate data were collected when librarians physically performed the tasks. Although in the beginning, this process is more time consuming, nonetheless, the final model considers real and detailed values about the library activities. Therefore, a trade-off between measurement time and accuracy must be considered. To document the activity flows, rounds of interviews with library managers and staff were done to identify the activities, resources and responsibilities. A second important insight is that software tools and the ease of presenting results help to decrease implementation time, and allow for better communication and validation. MS Office suite programs, such as Visio and Excel, were integrated to store, analyze and create graphical representations of the activity flows. As a consequence of this clear graphical representation, librarians were able to easily understand the sequences and their responsibilities in each process, and consequently, this allowed us to validate the collected information straightforwardly. Finally, a third important insight is that the involvement and commitment of the library staff are critical to the data collection in increasing the acceptance of the model. Therefore, motivation and an explanation of the measurement purpose are fundamental for achieving the desired commitment with staff. In the case of a large library, this requirement is even more critical since the number of employees gives rise to different types of opinions and attitudes regarding the process. This case study shows that TDABC is applicable to large libraries as well but that involvement of library staff is crucial.

A real situation that libraries face is the decision to automate repetitive processes. In this case study this is analyzed through three different situations of the returning process. With a simple but interesting example, we compare the costs of some specific activities performed by staff or machines. As we can see, the use of robots is well justified to automate these repetitive processes, especially in the cases of costly labor and in high number of activities.

This case study also illustrates some important benefits of TDABC such as the following: (1) better understanding of the costs' origins due to the disaggregated cost, time and resources per activity and of activities to be improved or discarded (e.g., including alerts in the LMS when a new request is available instead of printing requests); (2) an improved alternative evaluation for comparing different scenarios (e.g., manual activities vs. automated); (3) enhanced communication to analyze the cause of specific problems with stakeholders (librarians) that can easily understand the methodology applied; at the same time, librarians can justify the increase of wages or the development of the new services based on their responsibilities and the time required to perform them; (4) adaptability when, for instance, it is required to switch resources in busy periods, as when adding more staff to strengthen the user attention process is required to cope with the increased demand at the beginning of every semester; as demand increases (customers require extra attention to get familiar with the library), a shift from other areas for a specific period of time can be made; activities that were relegated for this cause can be prioritized during periods of low demand (e.g. when classes have ceased).

In summary, although at first glance, TDABC may seem more difficult to implement and to require more intensive data collection compared to a traditional costing system, our investigation shows that TDABC in practice is simple and easy to understand when the six steps identified by Everaert and colleagues (2008) are followed. Furthermore, the potential benefits accruing from the TDABC implementation such as the accuracy to calculate the costs of library services, the possibility of performing benchmarking analysis, disaggregating values per activity, and justifying decisions and

choices, validates the effort required to collect the data. An interesting avenue for future research is to perform a TDABC analysis on the user reference process. Considering that this kind of task does not follow a structured and systematic sequence (i.e., activities are addressed as they come in), as is in the case of our study, we expect that this analysis requires more effort and expertise. As a consequence, time analysis could be less accurate and more difficult to interpret. Additionally, a benchmark study is an interesting project to be done with libraries without fully automated loan and return processes.

References

- Blixrud, J. C. (2003). Assessing library performance: new measures, methods, and models. In *Proceedings of the IATUL Conferences* (p. Paper 9). Ankara, Turkey: Purdue e-Pubs. Retrieved from <http://docs.lib.purdue.edu/iatul/2003/papers/9/>
- Cardinaels, E., & Labro, E. (2008). On the Determinants of Measurement Error in Time-Driven Costing. *The Accounting Review*, 83(3), 735–756. <http://doi.org/10.2308/accr.2008.83.3.735>
- Ching, S. H., Leung, M. W., Fidow, M., & Huang, K. L. (2008). Allocating costs in the business operation of library consortium: The case study of Super e-Book Consortium. *Library Collections, Acquisitions, and Technical Services*, 32(2), 97–103. <http://doi.org/16/j.lcats.2008.08.013>
- Cooper, R., & Kaplan, R. S. (1988). Measure costs right: Make the right decision. *Harvard Business Review*, 66(5), 96–103.
- Dekeyser, R., & Holans, L. (2003). Around the World to Katholieke Universiteit Library, Leuven, Belgium. *Library Hi Tech News*, 20(8), 1.
- Ellis-Newman, J. (2003). Activity-Based Costing in user services of an academic library. *Library Trends*, 51(3), 333–348.
- Ellis-Newman, J., Izan, H., & Robinson, P. (1996). Costing support services in universities: An application of activity-based costing. *Journal of Institutional Research in Australasia*, 5(1), 75–86.
- Ellis-Newman, J., & Robinson, P. (1998). The cost of library services: Activity-based costing in an Australian academic library. *The Journal of Academic Librarianship*, 24(5), 373–379. [http://doi.org/16/S0099-1333\(98\)90074-X](http://doi.org/16/S0099-1333(98)90074-X)
- Everaert, P., & Bruggeman, W. (2007). Time-driven activity-based costing : exploring the underlying model. *Journal of Cost Management*, 21(2), 16–20.
- Everaert, P., Bruggeman, W., Sarens, G., Anderson, S. R., & Levant, Y. (2008). Cost modeling in logistics using time-driven ABC: Experiences from a wholesaler. *International Journal of Physical Distribution & Logistics Management*, 38(3), 172–191. <http://doi.org/10.1108/09600030810866977>
- Goddard, A., & Ooi, K. (1998). Activity-Based Costing and central overhead cost allocation in universities: A case study. *Public Money and Management*, 18(3), 31–38. <http://doi.org/10.1111/1467-9302.00124>
- Kaplan, R. S., & Anderson, S. R. (2004). Time-Driven Activity-Based Costing - Tool Kit. *Harvard Business Review*, (82), 131–138.
- Kaplan, R. S., & Anderson, S. R. (2007a). The innovation of time-driven activity-based costing. *Journal of Cost Management*, 21(2), 5–15.
- Kaplan, R. S., & Anderson, S. R. (2007b). *Time-Driven Activity-Based Costing: A simpler and more powerful path to higher profits*. Boston, MA, USA: Harvard Business School Press.
- Kaplan, R. S., & Cooper, R. (1998). *Cost & Effect: Using integrated cost systems to drive profitability and performance*. Boston, MA, USA: Harvard Business School Press.
- McKendrick, J. (2011). Funding and priorities: The library resource guide benchmark study on 2011 library spending plans (p. 40). Chatham, NJ, USA: Unisphere Research, a division of Information Today, Inc. Retrieved from <http://conan.lib.muohio.edu/ebooks/Fundingandpriorities.pdf>
- Novak, D. D., Paulos, A., & Clair, G. S. (2011). Data-driven budget reductions: A case study. *The Bottom Line: Managing Library Finances*, 24(1), 24–34. <http://doi.org/10.1108/08880451111142015>

- Novićević, B. M., & Ljilja, A. (1999). Total quality management and activity-based costing. *Facta Universitatis. Series: Philosophy and Sociology*, 1(7), 1–8.
- Pernot, E., Roodhooft, F., & Van den Abbeele, A. (2007). Time-Driven Activity-Based Costing for inter-library services: A case study in a university. *The Journal of Academic Librarianship*, 33(5), 551–560. <http://doi.org/16/j.acalib.2007.06.001>
- Rouse, W. B. (1975). Optimal resource allocation in library systems. *Journal of the American Society for Information Science*, 26(3), 157–165. <http://doi.org/10.1002/asi.4630260304>
- Siguenza-Guzman, L., Holans, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2013). Towards a holistic analysis tool to support decision-making in libraries. In *Proceedings of the IATUL Conferences* (p. Paper 29). Cape Town, South Africa: Purdue e-Pubs. Retrieved from <http://docs.lib.purdue.edu/iatul/2013/papers/29/>
- Stouthuysen, K., Swiggers, M., Reheul, A.-M., & Roodhooft, F. (2010). Time-Driven Activity-Based Costing for a library acquisition process: A case study in a Belgian University. *Library Collections, Acquisitions, and Technical Services*, 34(2-3), 83–91. <http://doi.org/16/j.lcats.2010.05.003>
- Sudarsan, P. K. (2006). A resource allocation model for university libraries in India. *Bottom Line: Managing Library Finances, The*, 19(3), 103–110. <http://doi.org/10.1108/08880450610682527>
- Tse, M. S. C., & Gong, M. Z. (2009). Recognition of idle resources in time-driven activity-based costing and resource consumption accounting models. *Journal of Applied Management Accounting Research*, 7(2), 41–54.
- Wegmann, G., & Nozile, S. (2009). The Activity-Based Costing method developments: State-of-the art and case study. *The IUP Journal of Accounting Research and Audit Practices*, 8(1), 7–22. <http://doi.org/oai:halshs.archives-ouvertes.fr:halshs-00482137>
- Yilmaz, R. (2008). Creating the profit focused organization using Time-Driven Activity Based Costing. In *EABR & TLC Conferences Proceedings* (p. 8). Salzburg, Austria: Clute Institute for Academic Research. Retrieved from <http://rifatyilmaz.net/SLZ08-191.pdf>

Chapter 5: **Time-Driven Activity-Based**

Costing Systems to maximize process

benchmarking in libraries

Siguenza-Guzman, L., Auquilla, A., Van den Abbeele, A., & Cattrysse, D. (2015). Using Time-Driven Activity-Based Costing to identify best practices in libraries. Submitted to the Journal of Academic Librarianship [Submission Being Reviewed].

This chapter provides more detailed insight of using TDABC to maximize process benchmarking in libraries. The chapter starts by describing the TDABC implementation in two academic libraries in Belgium. The workflow of ten library processes are compared and analyzed based on the TDABC results. Next, processes and performance improvements are reported and discussed in terms of cost and time, in particular with reference to the two libraries analyzed. This chapter ends by drawing implications and conclusions.

Apart from typographical adjustments, the content of this chapter is identical to the content of the published paper quoted above; where necessary, additional information or remarks are added in footnotes. The layout is adapted for consistency throughout this dissertation. Some redundancy with other chapters is unavoidable as an academic article needs its own introductory sections. This, however, entails the advantage that the chapter can be read separately.

Abstract

In the current competitive and dynamic environment, libraries must remain agile and flexible, as well as open to new ideas and ways of working. Based on a comparative case study of two academic libraries in Belgium, this research study investigates the opportunities of using Time-Driven Activity-Based Costing (TDABC) to benchmark library processes. To this end, we first start by describing the TDABC implementation. Then, we discuss and compare the workflow of ten library processes covering the four main library functions: acquisition, cataloging, circulation and document delivery. Next, we report and discuss potential processes and performance improvements that can be realized from using library time and costs information, in particular with reference to the two libraries analyzed. We conclude this article discussing the positive implications of using TDABC as a tool to enhance process benchmarking in libraries.

Contributions of the first author

Introduction outline, TDABC implementation at library 1, process standardization of both libraries, process benchmarking, analysis of results, and synthesis of the main discussions and conclusions.

5.1 Introduction

Over the last decades, libraries have been in a process of constant change: emerging digital services, high cost of information and continuing budget constraints have heightened the libraries need to improve their efficiency and their urgency to deliver high quality services at lower costs (Blixrud, 2003; ACRL Research Planning and Review Committee, 2010). In fact, libraries are passing through a challenging phase that forces them to retool their traditional services and resources. Among the approaches that can help libraries to improve their performance, benchmarking is considered as one of the most effective (Jean-Luc Maire, Vincent Bronet, & Maurice Pillet, 2005). Benchmarking can be very useful to libraries that are struggling with inefficient and uneconomical processes, and that are looking for more efficient ways to deliver their services (Henczel, 2002).

Benchmarking is the process of identifying, sharing and using local services, knowledge and practices, and then comparing against known best practices or the best in the field in order to determine and prioritize the areas that require improvement (Tardugno, DiPasquale, & Matthews, 2000; Jean-Luc Maire, 2002). This comparison can be executed internally when performances between institutional units are considered, or externally when different institutions' data are benchmarked. According to Forbes Gibb, Steven Buchanan, and Sameer Shah (2006), there are three main areas where benchmarking can be applied: performance, strategic and processes. Performance benchmarking relates to the comparison of outcomes or performance metrics among organizations, such as elements of price, speed, and reliability. Strategic benchmarking is focused on the understanding of strategic issues, on how successful enterprises are, and on the characteristics which contribute to or inhibit their success. Finally, process benchmarking uses process performance information to identify efficiency and effectiveness of processes and their corresponding workflows.

In particular, in the library sector, process benchmarking is used to compare the daily activities operations of libraries, and to determine existing differences and opportunities. This helps librarians to measure process workflows, as well as to ensure that libraries and their staff remain on the cutting edge of their profession. Process benchmarking is also useful to improve efficiency, effectiveness and competitiveness of libraries (Pauline Nicholas, 2010). In fact, libraries in many developed countries share statistical data regarding their processes and services in a regular basis. The results of this collaboration have been phenomenal as the strategic information gathered is used to demonstrate to top management that their performance is good or better than similar libraries, or conversely that they require a higher level of support from the mother institution to perform as well as others (Henczel, 2006).

Benchmarking studies, in general, utilize traditional metrics based on transactional aggregates, such as cost per set up, sales growth, and average order size (Anderson, 2006). However, as Steven R. Anderson (2006) indicates, the difficulty with these high level metrics is that they often fail to identify the true problem. Therefore, the combination of benchmarking analyses with internal understanding of performance drivers may allow for even greater efficiency and more accurate results. In this article we argue that Time-Driven Activity-Based Costing (TDABC) can be a useful tool to provide the internal understanding that benchmarking studies require. TDABC is a cost management technique developed by Kaplan and Anderson to overcome difficulties presented by previous costing systems such as traditional costing and activity-based costing methods (Kaplan & Anderson, 2007a; Siguenza-Guzman, Van den Abbeele, Vandewalle, Verhaaren, & Cattrysse, 2013). TDABC assigns resource costs directly to the cost objects using a fast and simple framework that only requires the unit cost of supplying resource capacity, and an estimation of the time required to perform an activity (Kaplan & Anderson, 2007b). The literature on TDABC outlines the following advantages: the ease and speed of building accurate costing models; the possibility of using multiple drivers; the good estimation of resource consumption and capacity utilization; the versatility and modularity to maintain and build inexpensive costing models; and the possibility of using TDABC in a predictive manner (Siguenza-Guzman et al., 2013). Besides these advantages and benefits, combining TDABC with other tools allows institutions for even greater improvement opportunities and results. Lorena Siguenza-Guzman, Alexandra Van den Abbeele, Joos Vandewalle,

Henri Verhaaren, and Dirk Cattrysse (2013), for example, summarize five possible combinations: 1) simulation modeling to analyze how to optimize resources since information is entirely composed of real values; 2) benchmarking tools to provide a deeper understanding of root problems such as sources of inefficiency and poor performance; 3) complementary information systems, such as bar codes, RFID technology and time sheets, to improve data accuracy and facilitate data collection; 4) balanced scorecard to facilitate translating strategy into performance measures, and to provide actionable performance measures for the balance scorecard; and 5) total quality management to help managers to identify non-value added activities.

By combining benchmarking with TDABC models, institutions can improve their performance “learning from others”, by means of comparing their processes, under the premise that most of these processes are common across multiple institutions (Kaplan and Anderson, 2007b). Besides, this combination allows comparing time equations and costs within different institution locations, such as departments and branches. Anderson (2006) analyzes this combination by illustrating how three companies in different sectors: distribution, banking and retail, use time-driven benchmarking models. According to this author, TDABC does not replace traditional benchmarking methodologies; rather, it enhances them. In fact, unlike traditional benchmarking that only reports macro results, TDABC isolates process differences to uncover root causes. In addition, a case study by Patricia Everaert, Werner Bruggeman, Gerrit Sarens, Steven R. Anderson, and Yves Levant (2008) in the logistic industry, shows how an internal benchmarking was performed in four warehouses to positively identify inefficiencies and synergy possibilities. Therefore, according to these previous studies, TDABC seems to offer a quick and inexpensive manner to improve benchmarking models by providing accurate and detailed information of sources of inefficiency and poor performance, as well as by helping to understand the impact that capacity utilization has on numbers.

In recent years, quite some research has been published on TDABC in libraries, but all these studies focus on specific library activities such as acquisition (Stouthuysen, Swiggers, Reheul, & Roodhooft, 2010; Kont, 2014), cataloging (Kont, 2013; Siguenza-Guzman, Van Den Abbeele, & Cattrysse 2014), circulation (Siguenza-Guzman, Van den Abbeele, Vandewalle, Verhaaren, & Cattrysse, 2014), and inter-library loan (ILL) (Pernot, Roodhooft, & Van den Abbeele, 2007). To our knowledge, no study has been published using TDABC to benchmark different library activities across several libraries. The aim of this article is therefore two-fold: (1) we analyze whether TDABC can be used to enhance process benchmarking in libraries, by means of the identification of “best practices” and opportunities for micro improvements; and (2) we provide more detailed insight on different library activities by using TDABC in academic libraries. The remainder of the article is as follows. Firstly, the different steps involved in implementing TDABC in two academic libraries in Belgium are explained. Secondly, the workflow of ten library processes (see Table 5.1) covering the four main library functions, namely, acquisition, cataloging, circulation and document delivery, are compared and analyzed based on the TDABC results. A special focus is given to them, as they are the main processes performed in most libraries, independent of their size or type. Next, processes and performance improvements that can be mirrored in the library time and costs are reported and discussed. The article ends by providing several conclusions drawn from the comparative study, and by discussing the implications of using TDABC as a Time-Driven model for process benchmarking in libraries.

5.2 Research methodology

For this comparative case study, we selected two Belgian academic libraries. Both libraries are dedicated to support education and scientific research in their corresponding universities. Library 1 is considered a medium-sized library that offers information sources on subjects of science, engineering and technology. This library offers extended opening hours for students, an electronic library that can be consulted at individual workplaces and facilities for guided self-education. Its services are handled by approximately 20.5 full-time equivalent employees (FTE). Library 2 is a small-sized faculty library, handled by approximately 8 FTE that serves the Faculties of Medicine,

Health and Pharmaceutical Sciences. Despite their differences in size, these two libraries were chosen because both libraries provide comparable services and have similar levels of automation.

Table 5.1: Processes to be analyzed using TDABC

Function	Process
Acquisition	Books acquisition
	Journals acquisition
Cataloging	Original cataloging
	Copy cataloging
Circulation	Lending items
	Returning items
Document Delivery	Requesting closed stack items
	ILL Outgoing Request
	ILL Incoming Request – digital items
	ILL Incoming Request – printed items

Data used in this study were collected through qualitative and quantitative methodologies. Qualitative interview data was combined with quantitative data analysis to evaluate the four main traditional library functions: acquisition, cataloging, circulation and document delivery. To calculate the cost of activities through TDABC, the six-step approach, proposed by Everaert et al. (2008), was followed. That is: 1) identifying resource groups; 2) estimating the total cost of each resource group; 3) estimating the practical time capacity of each resource group; 4) calculating the unit cost of each resource group; 5) determining the estimated time for each activity; and 6) multiplying the unit cost of each resource group by the estimated time for the activity.

To identify resource groups involved in each process, we conducted several rounds of interviews. We started the interviews with a discussion with the library manager, and then moved to a more detailed level by conducting interviews with all the employees of the library. Several resource costs were identified, including salaries, general overhead, subscription licenses, and equipment maintenance. After collecting this information in both libraries, ten common processes were identified and categorized based on their corresponding function as shown in Table 5.1.

Resource costs were first provided by accountants and managers who obtained data through the library information system (LIS). The costs were divided in two categories: direct and indirect. On the one hand, direct costs included salary of staff and student library employees - SLE (i.e. students hired to perform secondary activities), and maintenance cost of computers and radio frequency identification (RFID) systems. The latter refers to the self-scanning technology used to automate lending and returning processes. Besides the maintenance of RFID systems, it is important to consider the cost of the RFID tags that are the heart of a library RFID scheme. On the other hand, indirect costs included stationery, electricity, support, telephone, training, and other items used indirectly to perform an activity (Vazakidis & Karagiannis, 2009). At this point, our study revealed that both libraries had different resource costs such as wages and LIS subscription costs, complicating the process of benchmarking based on cost indicators. Therefore, an assumption was made for the sake of this study: the same resource costs were incurred in the two libraries. Although in principle this seems invalid for benchmarking purposes, the cost difference is still given based on the different resource types utilized. For instance, one library can be more efficient in terms of time because they employ highly qualified people and maybe the other library is less efficient because they use SLEs without enough experience. Consequently, original resource costs

were discarded and standardized by acquiring cost data from the Association of Research Libraries (ARL) statistics. ARL Statistics⁶ is a series of annual publications focused on describing and measuring the performance of ARL member libraries about their collections, expenditures, staffing and service activities. The cost data, such as salaries, general overhead and maintenance of technological equipment, were obtained from ARL statistics 2011-2012, and then converted from US dollar to euro equivalents at the exchange rate of that time period from x-rates.com (USD to EUR in 2011-2012=1.34). The period 2011-2012 was utilized since observations and interviews were performed in this time period.

The practical capacity of each resource group was then calculated by assuming an 80% of theoretical time capacity for people and 85% for machines (Kaplan & Anderson, 2007b). For staff capacity, 38 hours per week were accounted as theoretical time capacity. This results in a practical capacity of 30.4 hours per week (staff practical capacity = $38 * 80\%$), 1,824 minutes per day, 7,904 minutes per month, or 94,848 minutes per year. In the case of the machines, the theoretical time capacity was set equal to the time in which the libraries are open. Because library 1 and library 2 have different opening hours (69 and 63 hours respectively), the average time of 66 hours per week was defined as theoretical time capacity. Assuming 52 weeks per year, the practical capacity for machines is 175,032 min/year $\left(85\% \times 66 \frac{\text{hours}}{\text{week}} \times 60 \frac{\text{min}}{\text{hour}} \times 52 \frac{\text{weeks}}{\text{year}}\right)$.

Once the practical capacity was obtained, the cost per unit time was calculated by dividing the total cost of the resource by the practical capacity. An overview of the resulting costs is shown in Table 5.2.

Table 5.2: Costs involved in the analysis

Resource group	Cost per minute (€/min)
Librarian	0.59
Acquisition	0.71
Cataloging	0.72
Student library employee (SLE)	0.42
Library management system (LMS)	0.10
Computer maintenance	0.05
RFID labels	0.30
RFID maintenance	0.07
General overhead	0.10*

* General overhead is already incorporated to the resource groups corresponding to staff

Next, the required time to perform an activity was gathered through direct observation. Multiple observations were carried out for this data collection, using a stopwatch during several days at different hours in order to avoid possible biases (Siguenza-Guzman, Van den Abbeele, et al., 2014). Based on their average values, time equations were calculated for each activity. In order to standardize the procedures and time equations, activities were grouped and organized based on the core activities that any library performs in a process.

Eventually, costs were calculated for the activities by multiplying the unit cost per time of the resources by the estimated time required to perform the activity, and the total cost of each process was computed by summing up all activity costs.

⁶ <http://www.arl.org/stats/annualsurveys/arlstats/index.shtml>

5.3 Results

5.3.1 Acquisition

The acquisition department is responsible for selecting, requesting, ordering, purchasing and receiving new library materials and resources. It may also be in charge of budgeting and negotiating with publishers, dealers, or vendors, in order to obtain the items in the most economical and expeditious manner (Reitz, 2004). In acquisition processes, a distinction between books and journals is important because in some cases, these processes may involve different responsible actors, as well as different information may be entered into the system.

5.3.1.1 *Acquisition of books*

The process normally starts with a selection of new books performed by a specific person or department, such as the library director and the cataloging department. This acquisition order is received by the acquirer to be entered into the system. Next, the acquirer checks the request and prepares the order with basic information about the requested items. The acquisition order is sent to suppliers who in turn, send a reply confirming or denying the transaction. If the ordered books are available from suppliers, a package containing the invoice and the items is sent to the library. The acquirer receives the package, checks its content and completes the order in the system. Eventually, the invoice is sent to the financial department for its payment. The total costs of book acquisition in library 1 and 2 are presented in Table 5.3 and Table 5.4 respectively.

Tables 5.3–5.4 present the costs incurred and the time required to acquire a book in library 1 and library 2. The first column lists the standard activities identified in the process, while the second column indicates all the particular activities performed in the process by each library. The third column shows the average time per activity and the fourth column contains the cumulative time per standard activity. The fifth column indicates the accumulated costs of each resource group, the sixth column calculates the resulting cost incurred in the activity, and the last column shows the resulting cost per standard activity. Analysis of the resulting tables is provided in the discussion section.

5.3.1.2 *Acquisition of journals*

Similar to the above process, journal acquisition starts by selecting and then requesting new journals to suppliers. In contrast to book acquisition, the standard activities of selection, request, order and purchase occur only occasionally because libraries normally manage annual journal subscriptions. Nevertheless, libraries are constantly tracking and receiving journal issues. Table D.1 and Table D.2 in the Appendix D show the total costs and time incurred by libraries 1 and 2 respectively.

5.3.2 Cataloging

The cataloging department is responsible for producing bibliographical descriptions, subject analysis, and classification of books and other types of documents acquired by a library (Reitz, 2004). This unit is also responsible for physically preparing items for the shelves (Reitz, 2004). Cataloging processes can be divided in two categories: original and copy cataloging. The former refers to creating a new record from scratch, while the latter entails to adapting a pre-existing bibliographic record to the characteristics of the item in hand (Reitz, 2004). A standard cataloging process can be subdivided into 4 standard activities: searching, processing, labeling and shelving.

Table 5.3: Books acquisition process cost table (Library 1)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Selecting	Selecting books	0.95	1.60	0.86 ^a	0.82	1.37
	Check on the system	0.26		0.85 ^b	0.22	
	Check the bib description	0.39		0.85	0.33	
Requesting the book	Order with the bib description	0.52	2.09	0.85	0.44	1.78
	Put the order on the request	1.03		0.85	0.87	
	Prepare the order	0.33		0.85	0.28	
	Generate a request report	0.21		0.85	0.18	
Ordering	Make order to the suppliers	0.22	0.86	0.75 ^c	0.17	0.65
	Safe the orders online - folder	0.22		0.75	0.17	
	Read the confirmation	0.42		0.75	0.32	
Purchasing	Register invoice in LMS	1.01	4.00	0.85	0.86	3.17
	Pick up the invoice report	1.43		0.85	1.21	
	Put invoice in delivery room	1.56		0.70 ^d	1.10	
Receiving books	Check content of the package	1.85	4.85	0.70	1.30	3.61
	Register book in LMS	1.44		0.85	1.22	
	Put the book in delivery room	1.56		0.70	1.10	
<i>Total</i>			13.40			€ 10.57
^a 0.86 = <i>Cataloger + LMS + Computers</i> ^b 0.85 = <i>Acquirer + Computers + LMS</i> ^c 0.75 = <i>Acquirer + Computers</i> ^d 0.70 = <i>Acquirer</i>						

Table 5.4: Books acquisition process cost table (Library 2)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Selecting	Send acquisition order	1.00	1.33	1.79 ^a	1.79	2.03
	Check the order	0.33		0.70 ^b	0.23	
Requesting the book	Send e-mail order to bookstore	3.00	3.25	0.75 ^c	2.26	2.45
	Read the positive reply	0.25		0.75	0.19	
Ordering	Make the order	3.67	4.50	0.85 ^d	3.11	3.82
	Approve the order	0.83		0.85	0.71	
Purchasing	Receive book & invoice	0.75	5.00	0.70	0.53	3.86
	Register the invoice	2.42		0.85	2.05	
	Send invoice to Financial Department	1.83		0.70	1.28	
Receiving books	Check books	4.00	4.25	1.74 ^e	6.97	7.41
	Fill approval form	0.25		1.74	0.44	
<i>Total</i>			18.33			€ 19.57
^a 1.79 = <i>Director + Computers</i> ^b 0.70 = <i>Acquirer</i> ^c 0.75 = <i>Acquirer + Computers</i> ^d 0.85 = <i>Acquirer + Computers + LMS</i> ^e 1.74 = <i>Director</i>						

5.3.2.1 *Original Cataloging*

The process starts by searching the item on the library management system (LMS) in order to verify whether a similar record is already present in the database. If the item and record do not appear to match, the cataloger creates a new record which includes the bibliographic, holding and item description. Once the new record is stored in the database, the cataloger sticks labels to the item such as barcodes, RFIDs and stamps. Finally, the item is placed on the corresponding shelf or stack. The total costs of original cataloging processes in libraries 1 and 2 are presented in Table D.3 and Table D.4 respectively in Appendix D.

5.3.2.2 *Copy Cataloging*

In contrast to the above process, the cataloger finds a record that appears to match with the item. The cataloger validates and modifies the bibliographic description and then creates a new holding and item description. Labeling and shelving are similar to the original cataloging process. Table D.5 and Table D.6 show the total costs and time incurred in the copy cataloging process by libraries 1 and 2 respectively in Appendix D.

5.3.3 Circulation

The circulation unit is the service point where books and other materials are checked in and out of the library (Reitz, 2004). In addition to providing lending services, the circulation desk offers other specific services to users such as serials desk, ILL, basic search and reference services (Reitz, 2004). For the sake of process benchmarking, fines and reservation services are excluded as standard activities, since one of the two libraries does not offer such services (Budanov & Kavanosis, 2010).

5.3.3.1 *Lending Process*

A standard lending process can be subdivided in two main activities: searching and borrowing. A lending process starts with a user searching for an item on the library catalog. If an item appears to match with the information required, the user goes to the corresponding shelf and takes the item. The library user then goes to the circulation desk, places the item on the self-checkout machine, and performs the lending transaction. Finally, the library user prints a receipt which includes details of the borrowed items. The total costs of lending processes in libraries 1 and 2 are presented in Table D.7 and Table D.8 respectively in Appendix D.

5.3.3.2 *Returning Process*

In libraries, returning is the process of a user taking previously borrowed items back to the library. This process can be subdivided in three main activities: returning, classifying and shelving. The returning process is initiated by a user placing an item in the self-check machine and then printing the receipt as proof of returning. Next, the item is classified and re-shelved to the corresponding place. In Appendix D, Table D.9 and Table D.10 show the costs and time incurred in the returning process by libraries 1 and 2 respectively.

5.3.4 Document Delivery

Document delivery is the section responsible for the provision and delivery of books and other materials (physical and digital), usually for a fixed fee upon request (Reitz, 2004). The user normally requires picking up the printed material at the library, but electronic full-text may be

forwarded via e-mail. In this benchmarking study, three specific services are analyzed: requesting closed stack items, ILL outgoing request, and ILL incoming request.

5.3.4.1 *Requesting closed stack items*

Closed stack items are low-use titles and older library materials kept in storage areas inaccessible to users, in order to protect the collection or conserve space (Reitz, 2004). The process starts with a user searching on the library catalog and then requesting the item to circulation desk. Next, the librarian responsible collects the request, and goes to the closed stack to retrieve the item. The user receives the item from the circulation desk for its consultation. The process ends by returning the item to the circulation desk that in turn will re-shelve the item. A requesting closed stack item process can be subdivided into five main standard activities: searching, requesting, retrieving, delivering and shelving. In Appendix D, Table D.11 and Table D.12 show the total costs and time incurred in requesting closed stack items by libraries 1 and 2 respectively.

5.3.4.2 *ILL outgoing request*

Interlibrary loan is a service provided when items required by a user are unavailable in the library collection; users may request them from another library by filling out an interlibrary loan request (Reitz, 2004). ILL can be classified in two categories: outgoing requests and incoming requests (Pernot et al., 2007). The former refers to a library that requests and receives items from another library, while the latter refers to a library that receives requests and lends items to other libraries. ILL outgoing request can be subdivided in three main activities, namely ordering, receiving and delivering. The process is initiated by receiving and processing an ILL request from a user. The ILL responsible looks for a library possessing the requested item and orders it. Finally, the ILL responsible receives the item from the other library, and delivers the item to the user. Table D.13 and Table D.14 in the Appendix D show the cost table of the ILL outgoing requests by libraries 1 and 2 respectively. In this study, we focus on physical and digital journals since more than 90% of the ILL requests in both libraries are journals.

5.3.4.3 *ILL incoming request*

A standard ILL incoming request can be subdivided into four activities: requesting, retrieving, delivering and charging. Similar to outgoing request, ILL incoming request starts by receiving and processing an ILL request from another library. At this point, it is important to distinguish digital and printed articles because their delivery is substantially different. For a digital item, the ILL responsible searches the item online, sends the digital version to the customer and charges the transaction. Table D.15 and Table D.16 in Appendix D show the cost table of the ILL incoming requests for printed items by libraries 1 and 2 respectively.

Unlike digital materials, printed items are retrieved from the physical shelves, digitized and posted into the ILL system. Finally, the ILL responsible fills the price and closes the request as in the digital version. Table D.17 and Table D.18, provided in Appendix D, show the cost table of the ILL incoming requests for printed items by libraries 1 and 2 respectively.

5.3.5 Overview

Table 5.5 gives an overview of the benchmarking process that includes the total cost in euros and time in minutes of the different library processes of library 1 and library 2.

Based on the number of monthly repetitions, total time and costs per process are calculated. These monthly individual values are transformed into percentages of overall total time and cost

respectively in order to identify the processes that consume the majority of the library's time and resources cost. Figure 5.1 and Figure 5.2 illustrate the percentage of monthly time and cost consumed per process respectively.

Table 5.5: TDABC Cost Benchmarks between library 1 and library 2

Function	Process	Library 1		Library 2	
		Time (min)	Cost (€)	Time (min)	Cost (€)
Acquisition	Books acquisition	13.40	10.57	18.33	19.57
	Journals acquisition	13.01	10.13	20.47	18.59
Cataloging	Original cataloging	10.81	8.96	11.03	7.11
	Copy cataloging	7.89	6.46	10.23	6.43
Circulation	Lending items	1.85	0.24	1.18	0.14
	Returning items	1.11	0.31	1.94	0.74
Document Delivery	Requesting closed stack items	7.01	3.64	14.28	8.39
	ILL Outgoing Request	10.67	8.35	9.15	7.24
	ILL Incoming Request – digital items	4.05	3.13	3.92	3.18
	ILL Incoming Request – printed items	11.34	6.23	9.41	5.67

As can be seen in Figure 5.1, acquisition processes are among the three most time consuming activities in both libraries. The same trend is followed in the cost analysis (Figure 5.2) where these two processes are also among the most costly procedures. In library 1, the third place in both aspects: time and cost is occupied by original cataloging. On the contrary, the second and third place in library 2 with respect to time and cost respectively are occupied by the process of requesting closed stack items; original cataloging in this library is located in the last positions due to its little number of repetitions.

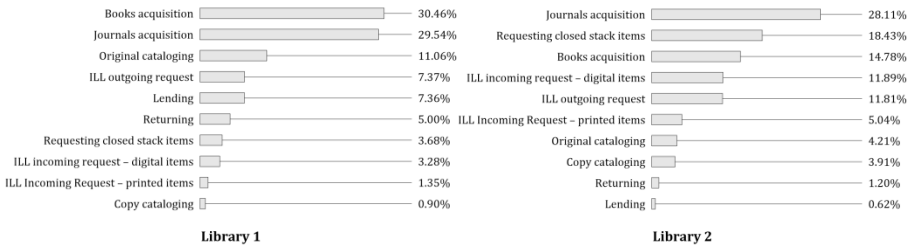


Figure 5.1: Percentage of time monthly consumed per process

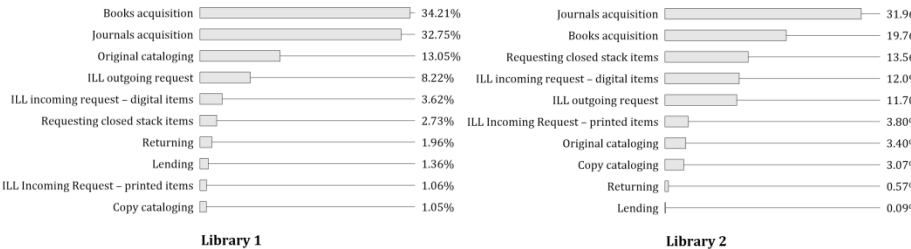


Figure 5.2: Percentage of cost monthly consumed per process

5.4 Discussion

A total of 10 processes were analyzed. On the basis of cost and time indicators, radar charts as shown in Figure 5.3(a) and Figure 5.3(b) are used to benchmark the different processes. These charts are very graphical in nature, making them easy to understand and capable of showing multiple dimensions simultaneously. In these two figures, two points are of interest: radar shape and library process comparison. Figure 5.3(a) shows the application of time comparison of the different processes. The original resulting time is standardized to a common interval scale with values between 0 and 1; therefore, a value close to "0" proportionally means a minor time consumed by a process, while a value close to "1" means a major time consumed by that process. Regarding the radar shape, library 1 displays a relatively balanced configuration. Both libraries perform equally well in terms of circulation processes and the ILL incoming request process for digital items, and underperform in terms of acquisition processes. This is not surprising since in both libraries, circulation processes are almost fully automated, while acquisition processes are mostly manual intensive. According to the process comparison, library 1 outperforms library 2 in 6 out of 10 processes, except in lending and ILL processes. An interesting point is the greatest divergence between library 1 and library 2 that can be seen with respect to the acquisition processes and the process of requesting closed stack items. Library 1 outperforms library 2 in more than 20% of three library processes: requesting closed stack items and acquisition of books and journals. The suboptimal performance of library 2 is the result of less efficient software, location of the closed stack collection, and a lack of specialized staff (more explanation is provided below). An additional highlight in this study is the time convergence of both libraries with respect to original cataloging, circulation processes and ILL processes. This is mainly due to the level of automation in circulation processes and the use of digital resources for the ILL services.

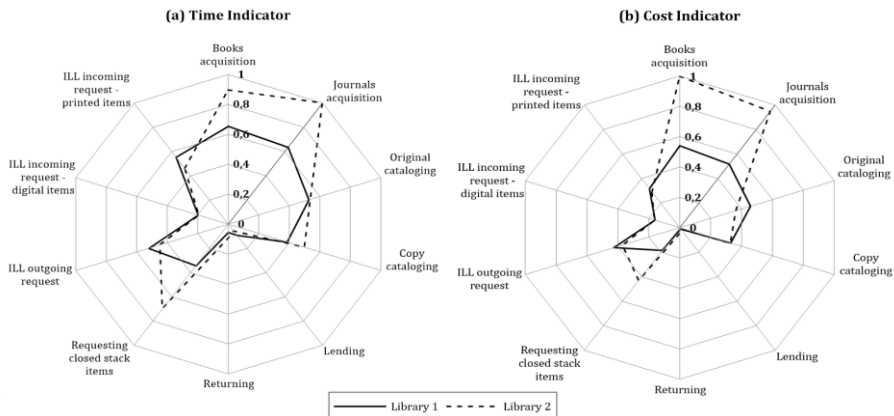


Figure 5.3: Process comparison based on time and cost indicators: (a) Time and (b) Cost

Figure 5.3(b) presents the same exercise, but for cost comparison. As can be seen in this figure, library 1 has the smallest overall surface, and thus has better performance than library 2. This is a consequence of library 1 outperforming library 2 in three specific processes: requesting closed stack items and especially in acquisition processes in more than 40%. One main aspect that contributes to the better results of library 1 in acquisition of books and journals in terms of cost efficiency is the minimal involvement of the library director in these activities⁷. On the contrary, an interesting point is the "dead heat" of these two libraries with respect to three processes, that is, copy cataloging, lending and ILL incoming request for digital items.

⁷ These activities are delegated to other academic responsible from various institutional departments.

Based on the radar analysis shown in Figure 5.3, three specific processes with major divergences in time and cost are selected to be deeply investigated. These are the acquisition process of books and journals, and the process of requesting closed stack items.

5.4.1 Book acquisition process

According to the results of the book acquisition process, costs in library 2 are 37% more costly than in library 1 and 85% more time-consuming. This mainly occurs, as shown in Figure 5.4, as result of a more elaborated and labored ordering and purchasing procedure. As can be seen in Table 5.3, Library 1 has the lowest unit cost, €10.57, to acquire one book. If library 2 with a unit cost of €19.57 for book acquisition (Table 5.4) wishes to be equally efficient in this service, it must reduce €9 to its unit cost. At first glance, library 1 may be seen as the model to follow in order to employ its best practices; however, this initial argument ignores specific details. For instance, although library 1 significantly outperforms library 2 in this process, two out of five activities are less time-consuming in library 2 because of a more simplified selecting and receiving procedure (one step less). Likewise, although the receiving books activity in library 1 is more time-consuming, the costs are more than double in library 2. This situation is due to the participation of the library director in these activities. Thus, this analysis just shows that both libraries have a lot to learn from each other.

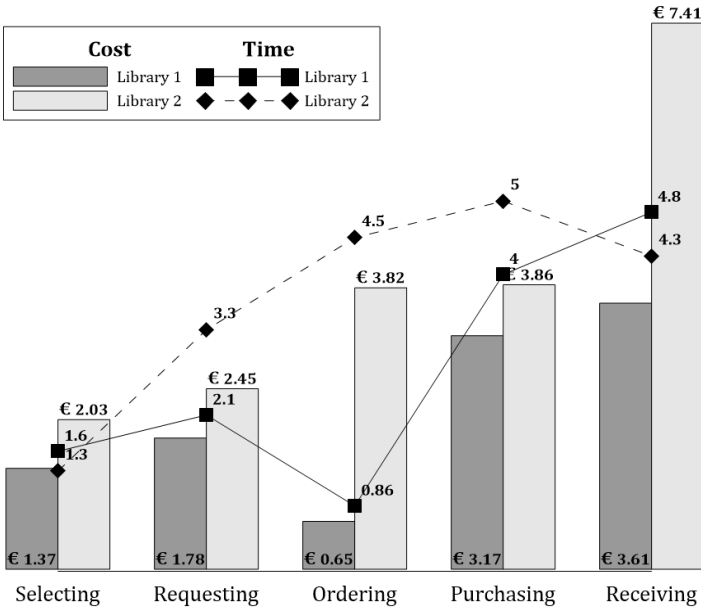


Figure 5.4: Book acquisition process comparison based on time and cost indicators

5.4.2 Journal acquisition process

Journal acquisition in library 2 is 83% more costly than library 1 and 57% more time-consuming. Even though library 2 is much more efficient in the purchasing activity (about 68%), a big difference can be observed in the receiving materials task, as shown in Figure 5.5, since library 2 requires nearly eight times more to register journal information in the LMS. This situation is caused by a more time-consuming software system and the lack of experience and expertise of the current responsible in these activities. Another important factor regarding time in library 2 is that it requires more logs and connections to the different library systems involved in the processes in

comparison to library 1. In addition, the library director of library 2 is involved in the receiving materials task, and in so doing, this task is dominant in terms of cost; this situation also holds for the books acquisition. In fact, in the book acquisition process, making an order in library 2 is much more complicated than in library 1; this situation makes the ordering activity to be the most time-consuming.

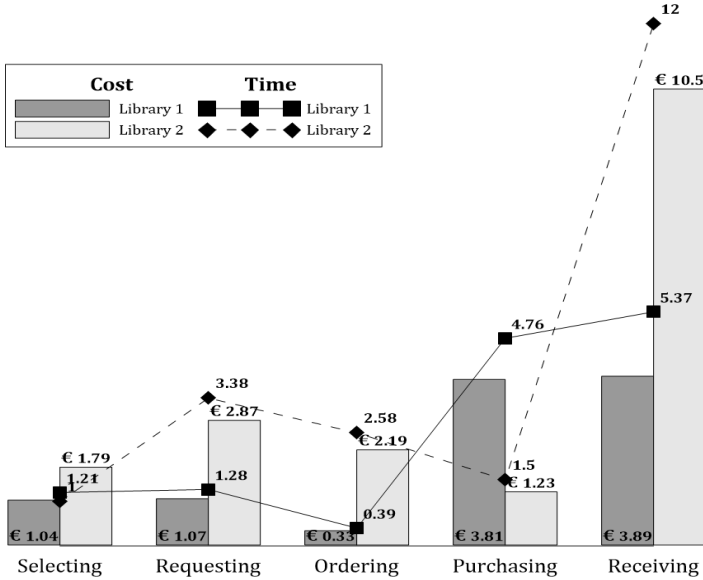


Figure 5.5: Journal acquisition process comparison based on time and cost indicators

5.4.3 Requesting closed stack items process

For the requesting closed stack items process, library 1 is clearly more efficient than library 2. Library 1 is twice efficient in terms of time and cost than library 2. The retrieving activity in library 2 represents more than half of the full time and cost (see Figure 5.6). Firstly, library 1, in the retrieving activity, performs the get item from the stack activity more than 4 minutes faster than library 2. This is mainly because in library 2, the closed stack collection is located about 40m away from the library entrance. Therefore, circulation staff spends a considerable amount of time going to this place, picking up the journal, and returning to the library. Moreover, library 2 has no employee strictly dedicated to closed stack activities as in library 1. Secondly, the shelving activity in library 1 is also performed 3 minutes faster than in library 2 due to the use of a sophisticated check-in system to pre-classify the items. Therefore, cheaper and simple attempts to reduce drastically this divergence can be made by incorporating batch practices in the logistic activities, such as retrieving and shelving.

On the contrary, just as the acquisition processes, even though library 1 seems to outperform significantly library 2, only two out of five activities are less costly and time-consuming in library 1. In fact, library 2 outperforms library 1 in both searching and requesting activities in more than 45% in terms of both time and cost. The principal reason of underperformance is primarily due to some non-value adding activities that library 1 has within this process, such as printing and handling the request form.

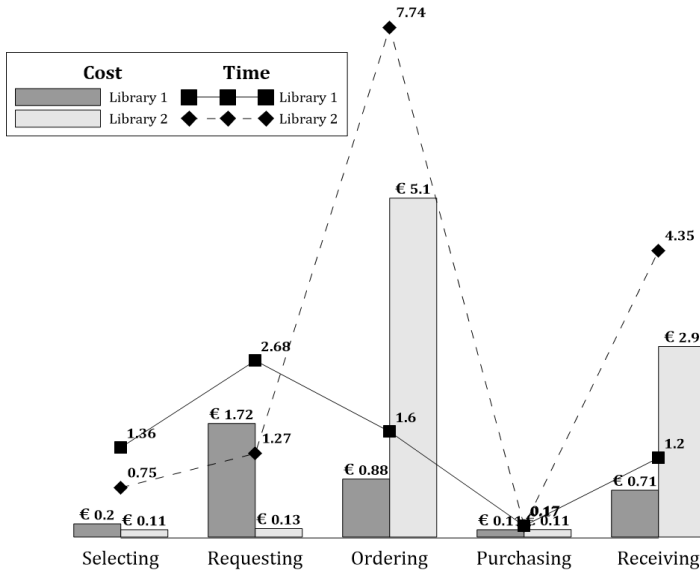


Figure 5.6: Requesting closed stack items process comparison based on time and cost indicators

5.5 Conclusions

As libraries readapt to meet the challenges of the current competitive and dynamic environment, the level of organizational complexity increases tremendously. This complexity comes in the form of retooling traditional services, creating new services, as well as shrinking budgets. In this challenging environment, measuring library performance cannot longer be done by looking only at overall analyses and outcomes. Benchmarking libraries provides real evidence that additional resources, technological and logistic changes, or support for infrastructure are needed. Internal benchmarking can be potentially used to better manage local processes by measuring and tracking their changes, to justify allocation and prioritization decisions, and to enable assessment activities.

This study has focused on developing a TDABC model for two Belgian libraries, with regard to four main library functions: acquisition, cataloging, circulation and document delivery. One of the most significant gains from this study is to evidence that TDABC makes available information about the costs of providing services, and disaggregates their corresponding causes. TDABC not only provides library managers with holistic information to make sounded decisions concerning the optimal resource allocation, but also provides managers with enough tools and strategic information to agilely identify improvement opportunities. In this article, the processes of library 1 and library 2 are compared in time and cost by means of TDABC analyses in order to highlight the best practices of both libraries. In the absence of TDABC analysis, the manager of library 1 can wrongly assume that in overall macro results, this library outperforms library 2 in all aspects, and that nothing needs to be changed in its processes. However, this study illustrates how both libraries have to learn from each other if wheels are not to be reinvented on both sides. Thus, mutually beneficial ways of improving library performance can be found through this type of comparison. Library 1, for instance, should focus on improving the scanning equipment for the ILL services, and eliminating the non-value-added steps coming from old ("legacy") procedures, such as printing and storing request forms. On the contrary, library 2 should focus on facilitating data entry into the LIS, relocating the closed stack collection, and delegating more responsibilities to low-wage employees like library assistants or students.

In addition, Time-Driven benchmarking provides strategic information to justify changes to superiors and staff in a good way and to make them much more aware of the outcomes and challenges that may occur during a change process. This benchmarking enhancement encourages paying more attention to reducing costs and trying to accomplish outcomes with fewer demands on library resources. It encourages rethinking roles, rules and activities across the library workflow without spending time on problems that have already been solved by exchanging knowhow among libraries. Time-Driven benchmarking helps to rethink how the time is spent within library processes, improve or streamline processes, reduce variability and standardize workflows.

Despite positive implications and results, some limitations of this study deserve consideration. First, although process improvements can be identified throughout comparative analysis, in some cases, certain aspects such as physical infrastructure and transportation distances cannot be easily changed or adapted. Second, even though both libraries provide comparable services and have similar levels of automation, each library may have different priorities. For instance, library 1 may emphasize quality in original cataloging, whereas library 2 focuses on fast copy cataloging or any other digital service.

References

- ACRL Research Planning and Review Committee. (2010). 2010 top ten trends in academic libraries. *College & Research Libraries News*, 71(6), 286–292.
- Anderson, S. R. (2006). Maximize Benchmarking with Time-Driven ABC: New Techniques that Change How We Measure Performance. *Acorn Systems*, 1–11.
- Blixrud, J. C. (2003). Assessing library performance: new measures, methods, and models. In *Proceedings of the IATUL Conferences* (p. Paper 9). Ankara, Turkey: Purdue e-Pubs. Retrieved from <http://docs.lib.purdue.edu/iatul/2003/papers/9/>
- Budanov, O., & Kavanozis, A. (2010, 2011). TD-ABC Applied for Libraries: Comparative study (Thesis submitted to obtain the degree of Master in Industrial Management). KU Leuven.
- Everaert, P., Bruggeman, W., Sarens, G., Anderson, S. R., & Levant, Y. (2008). Cost modeling in logistics using time-driven ABC: Experiences from a wholesaler. *International Journal of Physical Distribution & Logistics Management*, 38(3), 172–191. <http://doi.org/10.1108/09600030810866977>
- Gibb, F., Buchanan, S., & Shah, S. (2006). An integrated approach to process and service management. *International Journal of Information Management*, 26(1), 44–58. <http://doi.org/10.1016/j.ijinfomgt.2005.10.007>
- Henczel, S. (2002). Benchmarking--Measuring and Comparing for Continuous Improvement. *Information Outlook*, 6(7), 12–14, 17–18, 20.
- Henczel, S. (2006). Measuring and evaluating the library's contribution to organisational success. *Performance Measurement and Metrics*, 7(1), 7–16. <http://doi.org/10.1108/14678040610654819>
- Jean-Luc Maire. (2002). A model of characterization of the performance for a process of benchmarking. *Benchmarking: An International Journal*, 9(5), 506–520. <http://doi.org/10.1108/14635770210451491>
- Jean-Luc Maire, Vincent Bronet, & Maurice Pillet. (2005). A typology of "best practices" for a benchmarking process. *Benchmarking: An International Journal*, 12(1), 45–60. <http://doi.org/10.1108/14635770510582907>
- Kaplan, R. S., & Anderson, S. R. (2007a). The innovation of time-driven activity-based costing. *Journal of Cost Management*, 21(2), 5–15.
- Kaplan, R. S., & Anderson, S. R. (2007b). *Time-Driven Activity-Based Costing: A simpler and more powerful path to higher profits*. Boston, MA, USA: Harvard Business School Press.
- Kont, K.-R. (2013). Time-driven activity based costing for a library cataloguing process: A case study in Estonian University Libraries (pp. 74–76). Presented at the 10th Northumbria International Conference on Performance Measurement in Libraries and Information Services, York, UK: The University of York.
- Kont, K.-R. (2014). Using time-driven activity-based costing to support performance measurement in Estonian university libraries: A case study for acquisition process. In *Proceedings of*

- the IATUL Conferences (Vol. Paper 3). Espoo, Finland: Purdue e-Pubs. Retrieved from <http://docs.lib.purdue.edu/iatul/2014/acquisition/3>
- Pauline Nicholas. (2010). Benchmarking, an imperative for special libraries in the Caribbean: the Jamaican case. *Library Management*, 31(3), 186–197. <http://doi.org/10.1108/01435121011027354>
- Pernot, E., Roodhooft, F., & Van den Abbeele, A. (2007). Time-Driven Activity-Based Costing for inter-library services: A case study in a university. *The Journal of Academic Librarianship*, 33(5), 551–560. <http://doi.org/16/j.acalib.2007.06.001>
- Reitz, J. M. (2004). Dictionary for library and information science. Libraries Unlimited.
- Siguenza-Guzman, L., Auquilla, A., Van Den Abbeele, A., & Cattrysse, D. (2015). Using Time-Driven Activity-Based Costing to identify best practices in libraries: Costing Tables ([Secondary document]) (p. 10p.). KU Leuven.
- Siguenza-Guzman, L., Van Den Abbeele, A., & Cattrysse, D. (2014). Time-Driven Activity-Based Costing Systems for Cataloguing Processes: A Case Study. *LIBER Quarterly*, 23(3), 160–186.
- Siguenza-Guzman, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2013). Recent evolutions in costing systems: A literature review of Time-Driven Activity-Based Costing. *ReBEL - Review of Business and Economic Literature*, 58(1), 34–64.
- Siguenza-Guzman, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2014). Using Time-Driven Activity-Based Costing to support library management decisions: A case study for lending and returning processes. *Library Quarterly: Information, Community, Policy*, 84(1), 76–98. <http://doi.org/10.1086/674032>
- Stouthuysen, K., Swiggers, M., Reheul, A.-M., & Roodhooft, F. (2010). Time-Driven Activity-Based Costing for a library acquisition process: A case study in a Belgian University. *Library Collections, Acquisitions, and Technical Services*, 34(2-3), 83–91. <http://doi.org/16/j.lcats.2010.05.003>
- Tardugno, A. F., DiPasquale, T. R., & Matthews, R. E. (2000). *IT Services: Costs, Metrics, Benchmarking, and Marketing*. Prentice Hall Professional.
- Vazakidis, A., & Karagiannis, I. (2009). Activity-Based Management and traditional costing in tourist enterprises (A hotel implementation model). *Operational Research*, 11(2), 123–147. <http://doi.org/10.1007/s12351-009-0049-3>

PART III

Data Storage

Chapter 6: **Integrated Decision-Support**

System for library holistic evaluation

Siguenza-Guzman, L., Saquicela, V., & Catrysse, D. (2014). Design of an integrated Decision Support System for library holistic evaluation. Proceedings of IATUL Conferences. 35th Annual IATUL Conference. Helsinki, Finland, 2 – 5 June 2014 (pp. 1-12).

The third part of this dissertation, Data Storage, starts in this chapter with the analysis and design of an integrated decision support system based on data warehouse techniques for an academic library. The holistic approach described in Chapter 2 is used for data collection. Based on this mentioned approach, a set of queries of interest is described. Then, relevant data sources, formats, and connectivity requirements for a particular case study are identified. Next, data warehouse architecture is proposed to integrate, process, and store the collected data transparently. Eventually, the stored data are analyzed through reporting techniques, specifically on-line analytical processing tools.

Apart from typographical adjustments, the content of this chapter is identical to the content of the published paper quoted above; where necessary, additional information or remarks are added in footnotes. The layout is adapted for consistency throughout this dissertation. Some redundancy with other chapters is unavoidable as an academic article needs its own introductory sections. This, however, entails the advantage that the chapter can be read separately.

Abstract

The decision-making process in academic libraries is paramount, however, also highly complicated, due to the large number of data sources, processes, and high volumes of data to be analyzed. Academic libraries are accustomed to producing and gathering a vast amount of statistics about their collection and services. Typical data sources include integrated library systems, library portals and online catalogs, systems of consortiums, quality surveys, and university management. Unfortunately, these heterogeneous data sources are only partially used for decision-making processes due to the wide variety of formats, standards and technologies, as well as the lack of efficient methods of integration. This article presents the analysis and design of an integrated decision support system for an academic library. Firstly, a holistic approach documented in a previous study is used for data collection. This holistic approach incorporates key elements including process analysis, quality estimation, information relevance and user interaction that may influence a library's decision. Based on the mentioned approach above, this study defines a set of queries of interest to be issued against the integrated system proposed. Then, relevant data sources, formats and connectivity requirements for a particular example are identified. Next, data warehouse architecture is proposed to integrate, process, and store the collected data transparently. Eventually, the stored data are analyzed through reporting techniques such as on-line analytical processing tools. By doing so, the article provides the design of an integrated solution that assists library managers to make tactical decisions about the optimal use and leverage of their resources and services.

Contributions of the first author

The first author's contributions are: the introduction to the topic, research methodology, decision support architecture and conclusions.

6.1 Introduction

In a rapidly changing information environment characterized by the growing presence of e-content, emergence of new technologies, large amounts of data, and continuous diversification in user needs, knowledge management (KM) has become a powerful tool to promote innovation and to enable reengineering for library processes and services (ACRL Research Planning and Review Committee, 2010; Shanhong, 2002). At present, libraries can use KM as a way to expand their role in areas where they have had little impact, such as financial decisions and strategic decision-making (Townley, 2001). Although, the role of KM as a decision-making support tool has been well-documented in private sector organizations (Holsapple, 2001; Nicolas, 2004), its application in public sector institutions including universities, hospitals, and libraries remains immature (Tofan, Galster, & Avgeriou, 2013).

Knowledge-based Decision Support Systems (DSS) provide important information to analyze situations or conditions that impact operations and to make better and faster decisions (Poe, Brobst, & Klauer, 1997). In the case of libraries, several DSS have been documented; however, most of them mainly focus on specific areas such as budget allocation for physical and digital collections. Nevertheless, only a few studies are known to integrate other aspects such as: human resources, technological infrastructure, services, and library usage. The purpose of this article is to present the analysis and design of an integrated DSS (iDSS), which includes the aforementioned aspects based on Data Warehouse (DW) techniques for libraries. The remainder of the article is organized as follows. First, a description of the utilized holistic library analysis is briefly described. Then, the research methodology is outlined. Next, the design of an integrated DSS based on DW techniques is presented through a case study, and examples of the final result are reported. Finally, conclusions and future research directions are given in the last section.

6.2 Data collection through a holistic perspective

Implementing iDSS in libraries faces multiple challenges due to the high number of data sources, formats, and large volumes of data to be processed. In this context, Scott Nicholson (2004) proposes a theoretical framework that supports libraries gaining a thorough and holistic understanding of their users and services. Nicholson proposes a two-dimension matrix that evaluates libraries based on their library system and collection from internal and external perspectives. Due to the ease of understanding, completeness, as well as applicability to both physical and digital library resources, Lorena Siguenza-Guzman, Alexandra Van den Abbeele, Joos Vandewalle, Henri Verhaaren, and Dirk Cattryse (2015) adopted the framework as its basis to propose an architecture and an integrated set of tools to holistically assess libraries (see Figure 6.1). An example of this framework implementation is presented in a previous case study (Siguenza-Guzman, Holans, et al., 2013). By describing initial implementation stages, the authors probe the practical validity of the proposed holistic approach; however, the need of an iDSS to collect strategic information is strongly recommended.

	Library System	Library Collection
Internal Perspective (Library)	Service Analysis Processes, Time, Resources, Service Costs	Usage analysis Implicit and explicit data
External Perspective (Users)	Quality analysis Statistics gathering, Suggestion boxes, Usability testing, Satisfaction surveys	Collection analysis Citation patterns, Publishing patterns, Journals downloaded, and Journal's impact factor

Figure 6.1: Methodologies proposed for the economic evaluation of libraries through a holistic approach (Siguenza-Guzman et al., 2015)

The main characteristics of each quadrant and the methodologies proposed by Nicholson (2004) and Siguenza-Guzman et al. (2015) are the following.

1. *Internal perspective of the library system – Cost analysis processes*: costs and resources of library processes and services are analyzed. The authors describe three available methodologies: traditional costing system, activity-based costing system, and time-driven activity-based costing; recommending the last.
2. *External perspective of the library system – Quality*: quality of library processes and services are assessed by users. Siguenza-Guzman et al. recommend the use of at least one of the following methods: statistics gathering, suggestion boxes, Web usability testing, user interface usability, and satisfaction surveys.
3. *External perspective of the library collection – Bibliometrics*: the impact of the current library collection on its users is evaluated. The authors propose combining three methods: citation analysis, vendor-supplied statistics, and citation databases.
4. *Internal perspective of the library collection – Log analysis*: this quadrant analyzes usage patterns followed to manipulate the library system. Siguenza-Guzman et al. suggest the use of log analysis methods such as transaction and deep log analysis.

6.3 Research methodology

A successful approach for creating an iDSS based on DW techniques includes much more than the design process. Several decisions must be made, such as the DW data architecture to be used, data sources to be consulted, and the data integration scheme to be utilized. Thus, an adequate selection of methodology and technological tools for the construction of a DW will be instrumental in ensuring a successful implementation.

There are reasonably well-established approaches for implementing a DW; however, two classical methods are predominant: Inmon and Kimball. The Inmon methodology, or top-down approach, transfers the information from various Online Transactions Processing (OLTP) systems to a centralized DW, given that the DW has the following classic features: subject-oriented, integrated, time-variant, and nonvolatile (William H. Inmon, 2005). On the other hand, the Kimball methodology, or bottom-up approach, is the union of smaller data marts, where every data mart represents a business process or dimensional mode (Kimball, 2006). A data mart is a subset of the DW based on the same principles but with a more limited scope. After analyzing the Inmon and Kimball methodologies, a hybrid approach that integrates the best of both methodologies is adopted for this study. The DW methodology chosen for the case study implementation is Hefesto. The Hefesto methodology, created by Ricardo D. Bernabeu in 2007, starts by collecting information requirements and needs of the user, followed by the extraction of raw data, the transformation into standard formats, and the loading of the data into the DW database. The Hefesto methodology is characterized by the following features: it is easy, realistic, and simple to understand; it is based on user requirements gathering; it reduces the resistance to change; it uses conceptual and logical models, it can be applied to DW and data marts approaches, and it is independent of technologies, physical structure and life cycle type (Bernabeu, 2010).

Regarding the selection of technologies⁸, the market offers a wide range of software development products known as DW Business Intelligence (BI) tools. In this project, the Pentaho Community BI

⁸ The market offers a wide range of software development products known as Data Warehouse Business Intelligence (BI) tools, where companies like IBM, Google and Oracle have been leading and occupying the major section of the market (Chandrasekhar, Reddy, and Rath 2013). As the number of available BI tools continues to grow, the choice of the most suitable tool becomes increasingly difficult (Mikut and Reischl 2011). BI technologies consist of a suite of integrated applications available either as open-source or commercial software. In this work, Pentaho BI, an open-source tool, is selected to make the entire process of a Data Warehouse, which is as good as any expensive and closed source commercial software (Chandrasekhar, Reddy, and Rath 2013). Pentaho BI is an open-source integrated platform including data integration, ETL capabilities, data mining, reporting, OLAP

is selected to construct the DW based on the Hefesto methodology. The Pentaho BI⁹ is an integrated platform that includes: data integration, ETL (Extraction, Transformation and Load) capabilities, data mining, reporting, OLAP (on-line analytical processing) services and dashboard visualization.

6.4 Designing an integrated decision support system: Case Study

A case study to demonstrate the applicability of the holistic approach proposed to implement an iDSS based on DW was performed at the University of Cuenca (UC) library¹⁰. The UC library, or Regional Documentation Centre “Juan Bautista Vazquez”, is considered one of the most modern and biggest libraries in Ecuador. Its collection consists of about 250,000 books (i.e., 18 titles per student which is far above the national ratio), digital databases, and multimedia contents. The UC library, visited by an average of 1,200 students, is operated by 20 full-time staff members distributed in the main library and two branches.

6.4.1 Decision support system architecture

Based on the holistic approach proposed by Siguenza-Guzman et al. (2015), and the methodology and technologies selected to implement the DW, this article presents the resulting iDSS architecture implemented at the UC library as shown in Figure 6.2. The iDSS architecture of the UC library is structured in three layers: 1) *data sources* contains all sources used as data suppliers to the DW; 2) *data extraction, cleansing and storage* in charge of the design and implementation of ETL processes to maintain the DW; and 3) *data presentation* provides the appropriate reports for supporting information management and decision-making.

6.4.1 Hefesto methodology

This section describes the steps used to create the UC library DW based on the Hefesto methodology. This approach allows tackling the design of the DW from different detail levels, and reducing risks of failure and dissatisfaction by involving end-users early in the design process. The Hefesto methodology consists of the following four steps. 1) *Requirement Analysis* identifies the user information needs to define all queries of interest. 2) *OLTP analysis* is in charge of the data source analysis, determines how the indicators are built, defines correspondences and granularity, and builds the extended conceptual model. 3) *Logical Model* represents the structure of the DW; defines the type of implementation schema, the dimension and fact tables in order to create their respective unions. Facts are the core data elements being analyzed, while dimensions are attributes about facts. 4) *Data Integration* makes use of diverse tools such as cleansing techniques, data quality control, and ETL processes in order to integrate the data of different data sources; policies and strategies for the initial loading of the DW are also defined, as well as for its updating process. The DW creation according to the Hefesto structure is explained in more detail in the following subsections by way of examples.

services and dashboard visualization, which allows companies to develop problem-oriented, multidimensional solutions (Tuncer and van den Berg 2010). This platform is based on workflows, and process definitions that can be easily integrated. Pentaho suite is based on servers, engines, and third-party open source components, offering a scalable and sophisticated BI platform.

Keeping in line with the open source application language, the tool Waikato Environment for Knowledge Analysis – WEKA is used for the execution of the different data mining algorithms. WEKA is an open-source software, which consists of a collection of machine learning algorithms for data mining tasks. WEKA contains a comprehensive set of tools and algorithms for data processing, classification, regression, clustering, association rules and visualization (Frank et al. 2005).

⁹ Pentaho: <http://www.pentaho.com>

¹⁰ <http://biblioteca.ucuenca.edu.ec/>

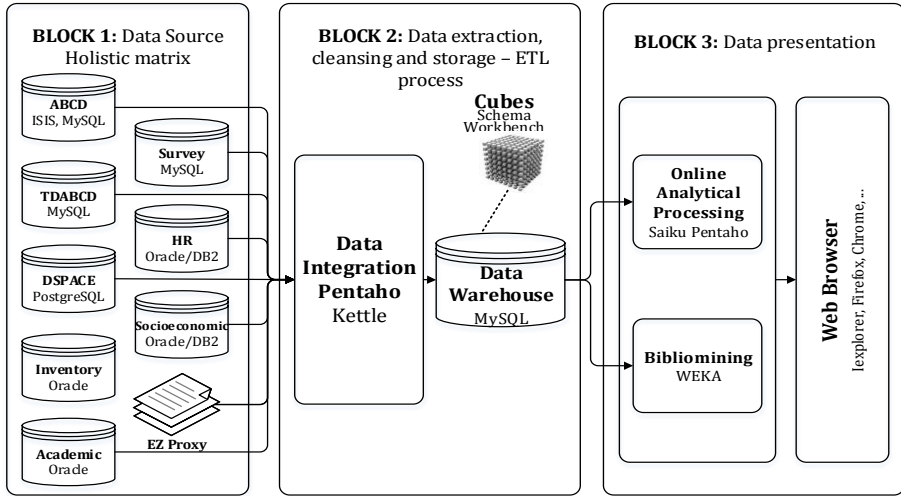


Figure 6.2: iDSS architecture of the UC library

6.4.1.1 BLOCK 1: Requirement Analysis

The requirements are analyzed through the four perspectives of the holistic evaluation framework proposed by Siguenza-Guzman et al. (2015) (Figure 6.1). Based on this structure, a list of queries (i.e. set of queries) of interest to be issued against the iDSS is defined. This list of requirements is collected through questions involving library needs. An example of the lending process requirements serves to exemplify how the questions are approached.

1. What are the number of loans of a particular item, of a given author and title, of a specific item category of a particular librarian in a time unit?
2. What are the number of loans of a particular librarian at a specific library branch in a time unit?
3. What are the operating costs of lending services at a particular campus in a time unit?
4. What are the number of returns of a particular librarian at a specific library branch at a given user type in a time unit?
5. What are the number of fines and its corresponding value in a time unit?

Once all questions are posed, the corresponding indicators and perspectives are identified. The resultant indicators and perspectives of the lending process example are shown in Figure 6.3.

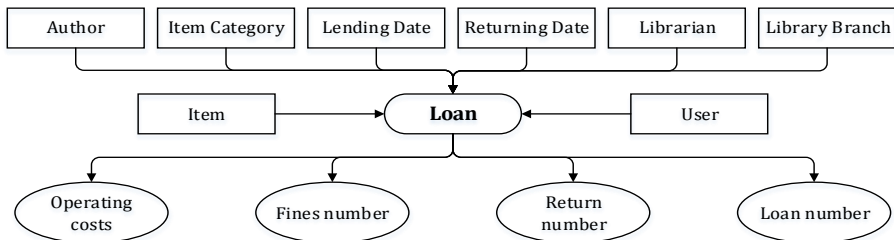


Figure 6.3: Indicators and perspectives of the lending process example

6.4.1.2 *OLTP Analysis*

The following step in the study is to identify the different data sources of the UC library based on the requirement analysis of the holistic evaluation approach. The holistic approach incorporates several key elements including process analysis, quality estimation, information relevance, and user interaction; thus, data have to be collected from internal and external sources. Internal data sources refer to the databases that are managed at the library level. On the contrary, external sources are not managed by the internal processes of the library. Collecting data from these heterogeneous sources presents a big challenge since generally different data sources use dissimilar formats and access methods including both structured (e.g., relational and documental databases) and unstructured data (e.g., word processing documents, spreadsheets, and log files).

After several meetings with the library manager, all the different OLTPs were documented. A total of ten data sources were identified and analyzed. These data are generated at library, university, and external levels. Library data sources are the following: ABCD, LibQUAL+®, DSPACE, and EZproxy. *ABCD*¹¹, a free and open-source integrated library automation software, offers the main functionalities of a library system such as acquisition, cataloging, circulation, online public access catalog (OPAC), and serials control. ABCD at the UC library uses MySQL and ISIS as its relational and documental database respectively, and MARC21 as its pre-existing cataloging structure. *LibQUAL+®*¹² is a proprietary set of services based on Web surveys that allows requesting, tracking, understanding, and acting upon users' perception of the quality of services offered by libraries. The LibQUAL+® survey consists of 22 questions, which are grouped into three quality dimensions: services provided, physical space, and information resources. *DSpace*¹³ is an open-source repository developed to provide access to digital resources. DSpace is implemented in Java and uses PostgreSQL as its database. *EZproxy*¹⁴ is proprietary software that allows libraries to offer their users remote access to the library e-sources. By default, events are recorded in standard web server log file format; however, EZproxy also includes the ability to add or remove fields to meet particular needs.

In addition, the UC library, as part of the university system, is linked with other university departments, implying information flows within the university. Data sources at the university level include the following: Olympe, GSocioeconomic, Academic, and HRM. *Olympe* is a commercial inventory manager software that uses Oracle as its relational database. *GSocioeconomic* is an in-house software tool responsible for the management of the socioeconomic data of university students. This software uses the Oracle database where all the student data are loaded. *Academic* is an in-house software tool that handles academic and enrollment process. This Web software uses Oracle as its relational database. *HRM* is an intranet-based application that manages data related to human resources. This in-house system uses Oracle as its database. Eventually, at an external level, other data are collected from sources such as Scopus reports and Ebsco statistics on online resources utilization. A summary of data sources, utilized to implement the DW at the UC library, are presented in Table 6.1.

6.4.1.3 *Logic Model*

Based on the conceptual data model (Figure 6.4), a data mapping from the OLTP sources to the logical model is performed. The logical model of the lending process is presented in Figure 6.5.

¹¹ <http://redde.bvsaude.org/projects/abcd>

¹² <http://www.libqual.org>

¹³ <http://www.dspace.org/>

¹⁴ www.ezproxy.com/

Table 6.1: Summary of data sources of the UC library

Application	Modules	Database	Type of database	Main attributes
<i>Internal data sources</i>				
ABCD	Cataloging	ISIS MARC21	Documental	Item ID, Author(s), Title, ..., Cataloger
	Circulation	MySQL	Relational	Lending, Returning, Fines, Reserves, Interlibrary loan
	TD-ABC-D	MySQL	Relational	Activities, Resources, Responsible, Processes, Times, Costs
	Reference Satisfaction Survey	MySQL	Relational	Librarian, Inquiry, Solution, Satisfaction level, PC Terminal, Survey
LibQUAL+®	Quality survey	MySQL	Relational	Campus, Questions, User, Time
Dspace	Institutional Repository	PostgreSQL Dublin Core	Relational	Community, Collection, Item, Bundle, Bitstream, Bitstream Format
EZproxy	Remote Access	Logs	LogFile	IP address host accessing, Date/time of request, URL Requested, Method of request, # bytes transferred
<i>University data sources</i>				
Olympos	Inventory	Oracle	Relational	Fixed assets, Class, Section, Area, Address, Provider, Fixed asset document
GSocioeconomic	Socioeconomic Data	Oracle	Relational	Entry/Exit students, Family members, Academic period, Student information, Faculties, Academic status
Academic	Enrollment	Oracle	Relational	Enrollment, Career, Academic period, Subject
HRM	Human Resources	Oracle	Relational	Personal information, User type, Department
<i>External data sources</i>				
Scopus reports	Citation database	Excel	Spreadsheet	Title, Publisher, 5-year Impact Factor, Topic
EBSCO statistics	Vendor-supplied statistics	Excel	Spreadsheet	Year, Month, Searches, Total full text, PDF full text, HTML full text, Image/Video, Abstract

In addition, the type of schema is also defined. Due to its simplicity and compatibility with the selected tools, the data schema that best fits the study is the “star” schema. In the star schema, facts are represented as a table in the center of the schema with multiple joins connecting it to the dimension tables. Thus, the dimension and fact tables are built to create their respective unions. As a result, multidimensional models are obtained. Figure 6.6 shows the multidimensional model of the lending process analysis, and consists of a central fact table and its respective dimensions.

6.4.1 BLOCK 2: Data Integration

After building the logical model, the relevant data generated by multiple sources are integrated by means of cleansing techniques, data quality control, and ETL processes; thus, allowing to have a clean and homogeneous version of the library data. Because this process is the most tedious and time-consuming part, literature recommends starting with a narrowly specific query and working through the entire process, and then, iteratively continuing developing the DW.

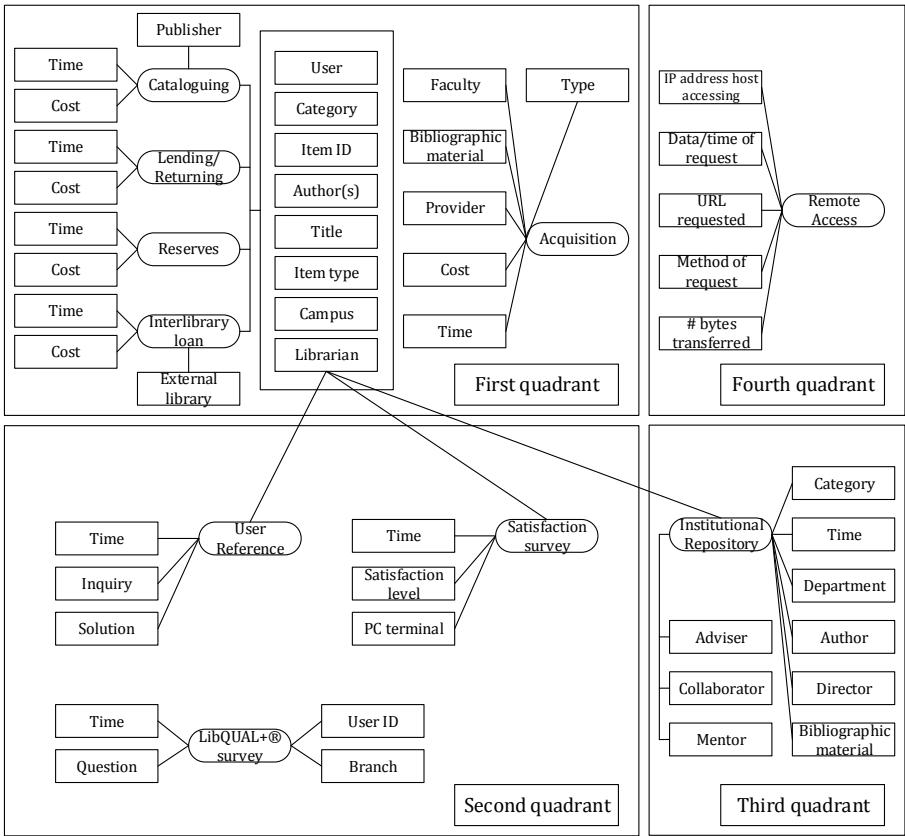


Figure 6.4: Conceptual data model of the UC library data warehouse

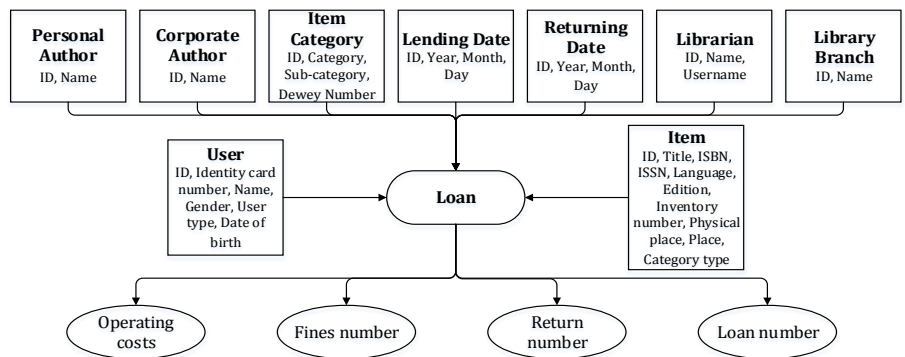


Figure 6.5: Logical model of the lending process example

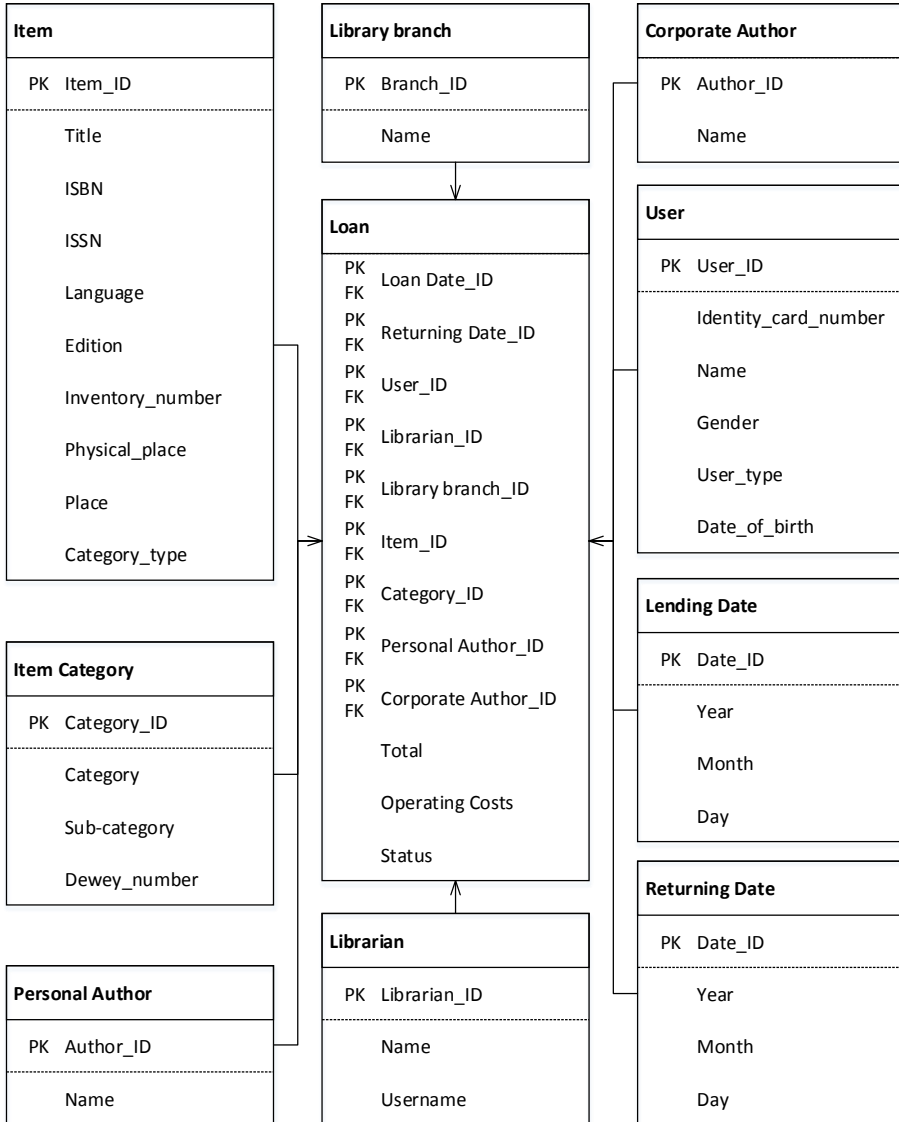


Figure 6.6: Multidimensional model of the lending process

6.4.1.1 *Extracting data*

Once the logical model is built, the following step is to extract the relevant data through ETL processes. In order to do so, the Kettle Pentaho suite is used. This tool includes a wide variety of components to access data sources such as relational databases, structured text files, and web services but lacks components to access metadata from documental databases as required in this case study. To solve this problem, a new component was developed in order to retrieve data from the ISIS database and to generate a structured text file with .mrc extension.

6.4.1.2 *Cleansing and transforming data*

Before loading data into the DW, these extracted data must go through a series of transformations in order to be cleaned. A particular example is the case of the data recorded through the ABCD system. Since the authority control module to establish uniform data entry is not used in the UC library, users easily make typing errors without any validation; for example, catalogers record data in the personal author field through different formats such as:

- *First surname, Names*
- *First surname Second surname, Names*
- *First surname Second surname, First name*
- *First surname, First name initial*

As shown in Figure 6.7, to solve the aforementioned problem, string similarity measures, such as the Jaro-Winkler metric, were used to indicate the percentage of similarity between fields. Finally, data according to the logic model are loaded in the different dimensions and fact tables.



#	Autor Personal	autor_personal	probabilidad
572	Agusti, M.	Aguirre S.	0,843560606
573	Agusti, M.	Aguirre S.	0,843560606
574	Agusti, M.	Agusti, Manuel	0,915151515
575	Agusti, M.	Agusti, Manuel	0,915151515
576	Agusti, M.	Aguirre S.	0,843560606
577	Agusti, M.	Aguirre S.	0,843560606
578	Agusti, M.	Agusti, Manuel	0,915151515
579	Agusti, M.	Agusti, Manuel	0,915151515
580	Agusti, M.	Aguirre S.	0,843560606
581	Agusti, M.	Aguirre S.	0,843560606
582	Agusti, M.	Agusti, Manuel	0,915151515
583	Agusti, M.	Agusti, Manuel	0,915151515

Figure 6.7: Evaluation of string similarity in the personal author field

6.4.1.3 *Loading data*

After extracting, cleansing, and transforming, data must be loaded into the warehouse. To do so, the Pentaho tool is used to create the multidimensional model in a relational database (MySQL). To optimize the model, indexes in the basic and commonly used searching fields are created.

6.4.2 **BLOCK 3: Data presentation**

Eventually, the stored data can be visualized and analyzed through reporting techniques located in the data presentation layer such as: data reporting, OLAP, and bibliomining tools. The tools utilized depend on the needs of the library manager to make decisions. In this study, OLAP tools, also called multidimensional analysis, are selected to produce reports and to be used by decision makers. OLAP tools can be used to prepare regular and unplanned reports, ensure quality, check data integrity, monitor the development of science, and evaluate or benchmark disciplines, fields or research groups (Hudomalj & Vidmar, 2003). According to the questions posed in the requirements analysis (Section 6.4.1.1), the results of the exploitation are:

1. Figure 6.8 answers the first question posed: What are the number of loans of a particular item, of a given author and title, of a specific item category of a particular librarian in a time unit?

Columnas

unidades

Filas

subcategoria

titulo

nombre_autor

username

Filtro

Info: 10:37 / 5 x 887 / 2.23s

subcategoria	titulo	nombre_autor	username	unidades
Literatura española y portuguesa	Algunas reformas para la enseñanza médica	ESPIHOZA V., VIRGILIO	lmontalvan	1
	América Latina se movilizaba contra el ALCA	REZENDE DE MELLO, ELSON	lmontalvan	1
			tbermeo	1
			retrospectiva	1
	Anales de la Universidad de Cuenca	BAQUERIZO MORENO, ALFREDO	rmendez	1
		CORDERO, LUIS	lcontreras	1
		FIGUEROA LORZA, JENNIE	lmontalvan	2
			rmendez	1
	Antología apócrifa	CORDERO Y LEÓN, RIGOBERTO	tbermeo	3
retrospectiva			3	

Figure 6.8: Number of loans of a particular item, of a given author and title, of a specific item category of a particular librarian

2. Figure 6.9 answers the second question posed: What are the number of loans of a particular librarian at a specific library branch in a time unit?

Columnas	unidades	
Filas	username	nombre_campus
Filtro		
username	nombre_campus	unidades
abcd	UC-CDJBV	32
acriollo	UC-CDJBV	10.872
	UC-SALUD	4
bcabrera	UC-YANUNCAY	912
dfajardo	UC-CDJBV	5.156
eduran	UC-CDJBV	2
	UC-YANUNCAY	1.251
fciriollo	UC-CDJBV	1.561
fruiiz	UC-CDJBV	48
gsegarra	UC-YANUNCAY	142
gmartinez	UC-CDJBV	9.562
	UC-SALUD	12
jmogollon	UC-CDJBV	4
	UC-YANUNCAY	1.617

Figure 6.9: Number of loans of a particular librarian of a specific library branch

3. Figure 6.10 answers the third question posed: What are the operating costs of lending services of a particular library branch in a time unit?

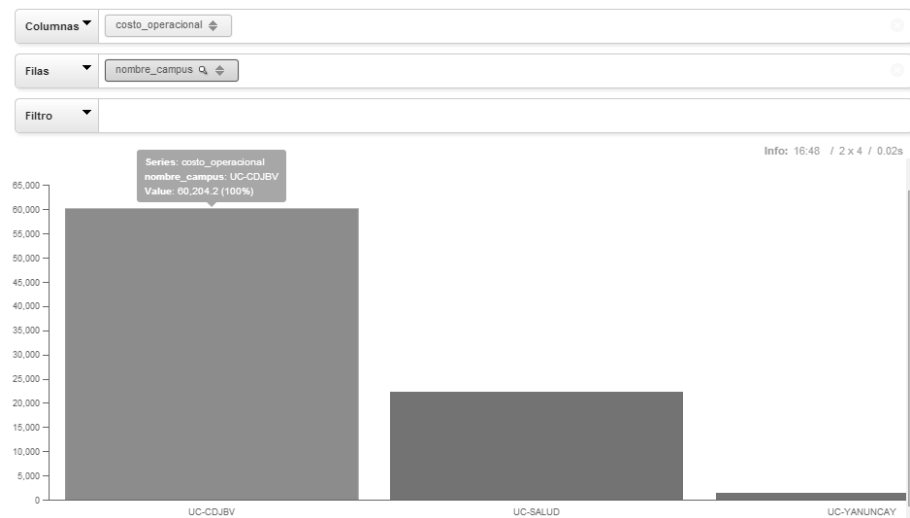


Figure 6.10: Operating costs of lending services of a particular campus

6.5 Conclusions

The main contribution of this work is the analysis and design of an iDSS for a university library through the case study analyzed. The distinguishing feature of the proposed architecture is the emphasis on the use of a holistic conceptual matrix to select the corresponding data sources. This decision implied integrating data from multiple and heterogeneous sources from the library, university, consortiums, and suppliers, all of which use dissimilar formats and access methods including both structured and unstructured data. Consequently, an adequate selection of methodology and technological tools for constructing the DW was necessary to ensure the data warehousing success. Important to note is that, thanks to the use of the Hefesto methodology at early deployment time, library managers and stakeholders were able to realize the potential of implementing an iDSS solution in order to make tactical decisions about the optimal use and leverage of their resources and services. Library managers can use this iDSS tool to ensure that different perspectives are taken into account in a decision-making process. In addition, the iDSS provides the data-based justifications for managerial and economic decisions library managers must make.

Work in progress includes the further refinement of the existing reports and the incorporation of additional sources to the integrated DSS, such as the syllabus management system, citation analysis, and Web portal statistics, so as to collect a wider range of information. Future work will focus on the analysis of information using bibliomining techniques, such as prediction and classification, in order to track patterns of behavior-based artifacts from library systems and thus, predict future library requirements. Furthermore, the plan to use semantic technologies to extend the multidimensional model is also proposed to enable the proper integration of knowledge in a way that is reusable by several applications across libraries.

References

- ACRL Research Planning and Review Committee. (2010). 2010 top ten trends in academic libraries. *College & Research Libraries News*, 71(6), 286–292.
- Bernabeu, R. D. (2010). Data Warehousing: Research and Concept Systematization - HEFESTO: Methodology for the Construction of a Data Warehouse. Cordova, Argentina.
- Chorba, R. W., & Bommer, M. R. W. (1983). Developing academic library decision support systems. *Journal of the American Society for Information Science*, 34(1), 40–50. doi:10.1002/asi.4630340107
- Hudomalj, E., & Vidmar, G. (2003). OLAP and bibliographic databases. *Scientometrics*, 58(3), 609–622. doi:10.1023/B:SCIE.0000006883.28709.d2
- Inmon, W. H. (2005). Building the data warehouse. John Wiley & Sons.
- Kimball, R. (2006). The data warehouse toolkit. John Wiley & Sons.
- Laitinen, M., & Saarti, J. (2012). A model for a library-management toolbox: Data warehousing as a tool for filtering and analyzing statistical information from multiple sources. *Library Management*, 33(4/5), 253–260. doi:10.1108/01435121211242290
- Nicholson, S. (2004). A conceptual framework for the holistic measurement and cumulative evaluation of library services. *Journal of Documentation*, 60(2), 164–182. doi:10.1108/00220410410522043
- Poe, V., Brobst, S., & Klauer, P. (1997). Building a Data Warehouse for Decision Support (2nd ed.). Upper Saddle River, NJ, USA: Prentice-Hall, Inc.
- Shanhong, T. (2002). Knowledge management in libraries in the 21st century. IFLA Publications - Libraries in the Information Society, 102, 88–93.
- Siguenza-Guzman, L., Holans, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2013). Towards a holistic analysis tool to support decision-making in libraries. In *Proceedings of the IATUL Conferences*. Cape Town, South Africa: Purdue e-Pubs.
- Siguenza-Guzman, L., Van Den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2015). A holistic approach to supporting academic libraries in resource allocation processes. *The Library Quarterly: Information, Community, Policy*, 85 (3).
- Townley, C. T. (2001). Knowledge Management and Academic Libraries. *College & Research Libraries*, 62(1), 44–55.

PART IV

Data Analysis and Presentation

Chapter 7: Literature review of data mining applications in libraries

Siguenza-Guzman, L., Saquicela, V., Avila-Ordóñez, E., Vandewalle, J., & Cattryse, D. (2015). Literature review of data mining applications in academic libraries. The Journal of Academic Librarianship, 41 (4), 499-510. <http://dx.doi.org/10.1016/j.acalib.2015.06.007>.

This chapter introduces data mining as an additional technique to analyze and visualize strategic library information. A comprehensive literature review and classification scheme for data mining techniques applied to libraries is provided. Forty-one empirical contributions over the period 1998-2014 are analyzed for their direct relevance. To do so, a detailed explanation of the research methodology adopted is first provided. This is followed by a description of the proposed method for classifying data mining applications in libraries. Classification results are then presented and discussed. The chapter finalizes by presenting limitations of the study, and by outlining research implications and prospects for future research developments.

Apart from typographical adjustments, the content of this chapter is identical to the content of the published paper quoted above; where necessary, additional information or remarks are added in footnotes. The layout is adapted for consistency throughout this dissertation. Some redundancy with other chapters is unavoidable as an academic article needs its own introductory sections. This, however, entails the advantage that the chapter can be read separately.

Abstract

This article provides a comprehensive literature review and classification method for data mining techniques applied to academic libraries. To achieve this, forty-one practical contributions over the period 1998-2014 were identified and reviewed for their direct relevance. Each article was categorized according to the main data mining functions: clustering, association, classification, and regression; and their application in the four main library aspects: services, quality, collection, and usage behavior. Findings indicate that both collection and usage behavior analyses have received most of the research attention, especially related to collection development and usability of websites and online services respectively. Furthermore, classification and regression models are the two most commonly used data mining functions applied in library settings.

Additionally, results indicate that the top 6 journals of articles published on the application of data mining techniques in academic libraries are: College and Research Libraries, Journal of Academic Librarianship, Information Processing and Management, Library Hi Tech, International Journal of Knowledge, Culture and Change Management, and The Electronic Library. Scopus is the multidisciplinary database that provides the best coverage of journal articles identified. To our knowledge, this study represents the first systematic, identifiable and comprehensive academic literature review of data mining techniques applied to academic libraries.

Contributions of the first author

Introduction to the topic, research methodology, classification method, verification of classification, analysis and tabulation of results, limitations, research implications and conclusions.

7.1 Introduction

Data mining, also known as knowledge discovery in databases, can be defined as the process of analyzing large information repositories and of discovering implicit, but potentially useful information (Han, Kamber, & Pei, 2011). Data mining has the capability to uncover hidden relationships and to reveal unknown patterns and trends by digging into large amounts of data (Sumathi & Sivanandam, 2006). The functions, or models, of data mining can be categorized according to the task performed: association, classification, clustering, and regression (Hui & Jha, 2000; Kao, Chang, & Lin, 2003; Nicholson, 2006b).

Data mining analysis is based normally on three techniques: classical statistics, artificial intelligence, and machine learning (Girija & Srivatsa, 2006). *Classical statistics* is mainly used for studying data, data relationships, as well as for dealing with numeric data in large databases (David J. Hand, 1998). Examples of classical statistics include regression analysis, cluster analysis, and discriminate analysis. *Artificial intelligence* (AI) applies “human-thought-like” processing to statistical problems (Girija & Srivatsa, 2006). AI uses several techniques such as genetic algorithms, fuzzy logic, and neural computing. Finally, *machine learning* is the combination of advanced statistical methods and AI heuristics, used for data analysis and knowledge discovery (Kononenko & Kukar, 2007). Machine learning uses several classes of techniques: neural networks, symbolic learning, genetic algorithms, and swarm optimization. Data mining benefits from these technologies, but differs from the objective pursued: extracting patterns, describing trends, and predicting behavior.

A typical data mining process, as shown in Figure 7.1, is an interactive sequence of steps that normally starts by integrating raw data from different data sources and formats. These raw data are cleansed in order to remove noise, and duplicated and inconsistent data (Han et al., 2011). These cleansed data are then transformed into appropriated formats that can be understood by other data mining tools, and filtration and aggregation techniques are applied to the data in order to extract summarized data. In fact, interesting knowledge is extracted from the transformed data. This information is analyzed in order to identify the truly interesting patterns. Eventually, knowledge is visualized to the user. More detailed information regarding a data mining process can be found in Han et al. (2011).

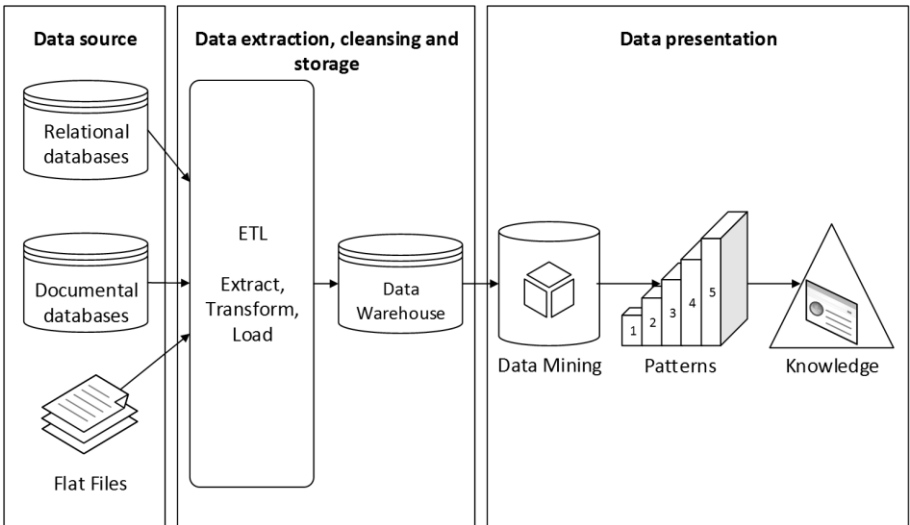


Figure 7.1: Data mining process, based on Han et al. (2011).

Data mining techniques are applied in a wide range of domains where large amounts of data are available for the identification of unknown or hidden information. In this sense, N. Girija and S.K. Srivatsa (2006) indicate that data mining techniques used in *www* are called web mining, used in text are called text mining, and used in libraries are called bibliomining.

The term bibliomining, or data mining for libraries, was first used by Scott Nicholson and Jeffrey Stanton (2003) to describe the combination of data warehousing, data mining and bibliometrics. This term is used to track patterns, behavior changes, and trends of library systems transactions. Although the concept is not new, the term bibliomining was created to facilitate the search of the terms “library” and “data mining” in the context of libraries rather than in software libraries. Bibliomining is an important tool to discover useful library information in historical data to support decision-making (Kao et al., 2003). However, to provide a complete report of the library system, bibliomining needs to be used iteratively applied in combination with other measurement and evaluation methods; as strategic information is discovered, more questions may be raised and thus start the process again (Nicholson, 2003b).

Bibliomining, as any knowledge extraction method, needs to follow a systematic procedure in order to allow an appropriate knowledge discovery. The bibliomining process starts by determining areas of focus and collecting data from internal and external sources (Nicholson, 2003b). Then, these data are collected, cleansed, and anonymized into a data warehouse. To discover meaningful patterns in the collected data, the bibliomining process includes the selection of appropriate analysis tools and techniques from statistics, data mining, and bibliometrics (Nicholson, 2006a). Interesting patterns are analyzed and visualized through reports. The mining process will be iterated until the resulted information is verified and proved by key users such as librarians and library managers (Shieh, 2010).

The application of bibliomining tools is an emerging trend that can be used to understand patterns of behavior among library users and staff, and patterns of information resource use throughout the library (Nicholson & Stanton, 2006). Bibliomining is highly recommended to provide useful and necessary information for library management requirements, focusing on the professional librarianship issues, but highly database technical dependent (Shieh, 2010). Bibliomining can also be used to provide a comprehensive overview of the library workflow in order to monitor staff performance, determine areas of deficiency, and predict future user requirements (Prakash, Chand, & Gohel, 2004). The resulting information gives the possibility to perform scenario analysis of the library system, where different situations that need to be taken into account during a decision-making process are evaluated (Nicholson, 2006a). An additional application is to standardize structures and reports in order to share data warehouses among groups of libraries, allowing libraries to benchmark their information (Nicholson, 2006a). Therefore, in order to improve the interaction quality between a library and its users, the application of data mining tools in libraries is worth pursuing (Chang & Chen, 2006).

The aim of this study is to investigate how far academic libraries are pragmatically using data mining tools, and in which library aspects librarians are implementing them. To this end, content and statistical analyses are used to examine articles that include case studies of academic libraries implementing data mining tools. The remainder of the article provides a detailed explanation of the research methodology adopted in this literature study. This is followed by a description of the proposed method for classifying data mining applications in libraries. Classification results are then presented and discussed. The article concludes by presenting limitations of the study, and by outlining research implications and prospects for future research.

7.2 Research methodology

The present study follows the methodology employed by E.W.T Ngai, Karen K. L. Moon, Frederick J. Riggins, and Candace Y. Yi. (2009) to analyze and classify data mining techniques applied to customer relationship management. In this study, the analysis and classification are based on the examination of selected search engines and the use of a set of descriptors, all related to their specific interests. Then, the selected articles are reviewed and categorized based on a classification

framework. The resulting list and classification is independently verified by research triangulation; finally, findings are reported in order to identify implications and future research directions.

Thus, following the Ngai et al. selection criteria and evaluation framework, a Web-based literature research on practical documents about data mining applications was conducted in order to identify relevant articles. As the nature of research on data mining and libraries is difficult to comprehend within the confines of specific disciplines, the relevant articles are scattered throughout numerous scholarly journals. Consequently, bearing in mind the degree of relevance or specialization to the subject analyzed, a set of four search engines was first selected to perform journal browsing. Based on the specialization degree, two major Library and Information Science (LIS) databases were searched: Library Information Science & Technology Abstracts (LISTA) accessed through EBSCOhost, and Library and Information Science Abstracts (LISA) accessed through ProQuest. In addition, two multidisciplinary databases: Web of Science (WoS) and Scopus were also consulted as complementary databases, as both search engines are among the largest and most common of the multidisciplinary databases available. Subsequently, citation tracing was also employed to discover additional papers relevant to this study; thus, the reference section for each article found was traced in order to find additional journal articles.

The search was operated according to the following procedure. First, a selection of subject terms was performed in order to identify terms that represent the concepts related to the topic under the study. To this end, the thesauruses of LISA and LISTA were consulted to draw up a set of standardized descriptors. Although the term “bibliomining” does not appear in both thesauruses, it was also incorporated as a subject term in order to investigate if academics and practitioners utilize this word as part of their titles or provided keywords. Based on this selection of terms, relevant articles were searched by combining the following subject terms: “data mining”, “academic librar*” and “university librar*”. The asterisk (*) is used to find words ending with a common stem, for example, librar* = libraries or library. All these search terms and their combinations were searched in subject headings (article title, abstract and keywords), and the analysis was limited to journal articles published in English. An overview of the criteria and results is shown in Table 7.1. When the number of articles searched for was within a reasonable number to conduct analysis, the resultant literature was sorted, summarized, and discussed in order to generate a final sample consisting of 485 potentially relevant studies. Then, the full text of each article was retrieved for detailed evaluation in order to eliminate those articles that did not meet the selection criteria with the application of data mining techniques in academic libraries. Each excluded article was registered in an excluded-studies table, followed by an explanation for its separation. All excluded articles were further screened by a different reviewer to confirm agreement with exclusion. In addition, the reference section for each included article was examined for possible titles of additional studies. By so doing, a total of 135 extra articles were analyzed. The standardized inclusion/exclusion criteria were as follows:

- Only English articles were included in the study.
- Only the articles related to the application of data mining techniques in academic libraries were selected, as these were the focus of this literature review.
- The articles describing the application of data mining techniques in academic libraries without a specific case study were excluded.
- Only the articles clearly describing how the mentioned data mining technique(s) could be applied and assisted in library settings were selected.
- Masters and doctoral dissertations, conference papers, text books and unpublished working papers were excluded. The main reason for this decision was that both academics and practitioners most commonly use journals both to acquire information and spread new knowledge (Gonzalez, Llopis, & Gasco, 2013). Whereas journal articles currently represent the highest level of research, other formats, like books, are confined to gathering and spreading knowledge that is already established. As for conferences, it is usual for most valuable articles to end up being published in journals; in fact, the conference represents a step prior to the definitive journal publication (Gonzalez et al., 2013).

Table 7.1: Search criteria and number of results per database

Database	Query	Number of results
LISTA	Bibliomining	11
	"data mining" "academic librar*"	20
	"data mining" "university librar*"	16
LISA	Bibliomining	9
	"data mining" "academic librar*"	9
	"data mining" "university librar*"	11
Web of Science	Bibliomining	14
	"data mining" "academic librar*"	45
	"data mining" "university librar*"	12
Scopus	bibliomining	31
	"data mining" "academic librar*"	140
	"data mining" "academic librar*" "case stud*"	53
	"data mining" "university librar*"	92
	"data mining" "university librar*" "case stud*"	22
Total articles analyzed		485

Forty-one articles were subsequently selected. A detailed table of the selected articles can be found in Appendix E. Each selected article was carefully reviewed and separately classified according to four quadrants of a holistic evaluation matrix for libraries and four main data mining functions, as shown in Figure 7.2. Although this research was not exhaustive and oriented to the application of data mining techniques in academic/university libraries, it serves as a comprehensive base for an understanding of data mining research in libraries in general.

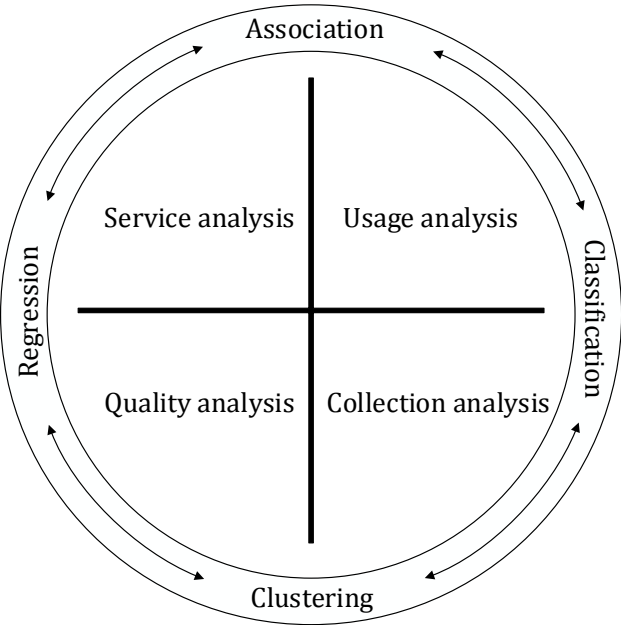


Figure 7.2: Classification framework for data mining techniques based on the Ngai et al. (2009) approach

7.3 Classification method

7.3.1 Classification framework: Holistic approach for library evaluation

Facing rapidly changing landscapes, characterized by shrinking budgets and dynamic services, libraries have recognized the need for evidence of their value. Academic libraries more than ever before, are called upon to demonstrate and justify their existence, and their contribution to institutional missions and goals (Association of College and Research Libraries, 2010). In fact, new trends and issues affecting academic libraries include a culture of increasing accountability for outcomes, in which libraries will be required to find better ways to document these connections (ACRL Research Planning and Review Committee, 2014). Nicholson (2004) and Siguenza-Guzman et al. (2015) recognizing the need to evaluate libraries in a holistic and structured manner, propose the use of a two-dimensional evaluation matrix. The four quadrants of the holistic evaluation matrix are the following:

1. *Internal perspective of the library system – Process/service analysis:* In this quadrant, the “library system” refers to everything that is part of the offerings of the library such as the organizational scheme, electronic equipment, library staff, and facilities. Internal perspective of the library system involves analyzing the topics related to processes and services carried out within the library.
2. *External perspective of the library system – Quality analysis:* Quality of the collection and services are assessed by users. Thus, the second quadrant evaluates the aboutness, pertinence, and usability of physical and digital resources by exploring users’ perceptions (Nicholson, 2004). Assessment methods to measure the quality of services and collection include statistics gathering, suggestion boxes, Web usability testing, user interface usability, and satisfaction surveys (Wright & White, 2007).
3. *Internal perspective of the library collection – Collection analysis:* The third quadrant aims to evaluate the usefulness of the library collection. Proponents of this holistic approach suggest the combination of three assessment methods; namely, citation analysis, vendor-supplied statistics, and citation databases. By doing so, libraries will gain an extensive knowledge about their collection value and information relevance.
4. *External perspective of the library collection – Usage analysis:* This final quadrant evaluates users’ behavior when manipulating the library system. Users’ interaction with the system is utilized to study users’ preferences to personalize library services. Transaction log analysis, Web usage analysis, deep log analysis, and usage statistics are the main techniques utilized for this purpose.

Each quadrant of this evaluation matrix shares the common goal of supporting libraries in gaining a thorough and holistic understanding of their users and services. Data mining techniques, therefore, can help to accomplish such a goal by uncovering hidden patterns of behavior among library users and staff members, and patterns of information resource usage (Nicholson & Stanton, 2003).

7.3.2 Classification framework: Data mining models

Bibliomining can reveal issues associated with information-seeking user behavior, predict future trends on collection development, and build user communities based on common information interests. Based on the type of knowledge discovery, data mining functions can be divided into unsupervised and supervised algorithms (Chen & Liu, 2004). The former recognizes relationships in non-classified data, while the latter requires the data to be pre-classified in order to explain those relationships. According to these two main function types, data mining algorithms can be divided into the following categories: association, clustering, classification, and regression (Fayyad, Piatetsky-Shapiro, & Smyth, 1996; D. J. Hand, Mannila, & Smyth, 2001).

1. *Association*: The so-called association rule aims to find the existing (or potential) relationships between data items in a database such as attributes and variables (Lunfeng, Huan, & Li, 2012). Examples of common association tools are statistics and apriori algorithms (Ngai et al., 2009).
2. *Clustering*: Clustering is the task of uncovering unanticipated trends by segmenting no predefined clusters. This approach is used in situations where a training set of pre-classified records is unavailable (Chen & Liu, 2004). Common tools for clustering include neural networks, k-means algorithms, and discrimination analysis (Ngai et al., 2009).
3. *Classification*: Classification is the task of attempting to discover predictive patterns by classifying database records into a number of predefined categorical classes based on certain criteria (Chen & Liu, 2004). Common classification tools are neural networks, decision trees, and if-then-else rules (Ngai et al., 2009).
4. *Regression*: Regression is an essentially statistical technique that maps a data item to a real-valued prediction variable. This data mining function is normally used to capture the trends of frequent patterns. Examples of common regression techniques include linear regression and logistic regression analysis.

In turn, numerous data mining techniques are available for each type of data mining function. The choice of the data mining technique depends on the nature and purpose for the research study or the library requirements (Banerjee, 1998). Examples of some widely used data mining algorithms include the following: k-means algorithms for clustering, association rules for association, linear and logistic regression for regression, and decision trees for classification.

7.3.3 Research classification process

Following the Ngai et al. (2009) classification process approach, each selected article was reviewed and classified according to the proposed classification framework by three independent researchers. Researcher A was selected based on his/her expertise on the library holistic approach, whereas researchers B and C were selected based on their data mining experience. The classification process consists of five phases:

1. Online database search.
2. First classification by researcher B.
3. First verification of classification results and excluded articles by researcher A.
4. Independent verification of classification results by researcher C.
5. Discussion on classification results by researchers A, B and C, and
6. Analysis and tabulation of results by researchers A and B.

If a discrepancy in classification results existed between researchers, each article was then discussed until an agreement was reached by consensus on how the article should be classified from the final set in the proposed classification framework. Figure 7.3 shows the selection process utilized across the study. The collection of articles was analyzed based on the library holistic matrix and data mining models, by year of publication, country of implementation, journal in which the article was published, as well as the type of library analyzed (physical, digital or both).

7.4 Classification of the articles

A detailed distribution of the 41 articles classified by means of the proposed framework is shown in Table 7.2.

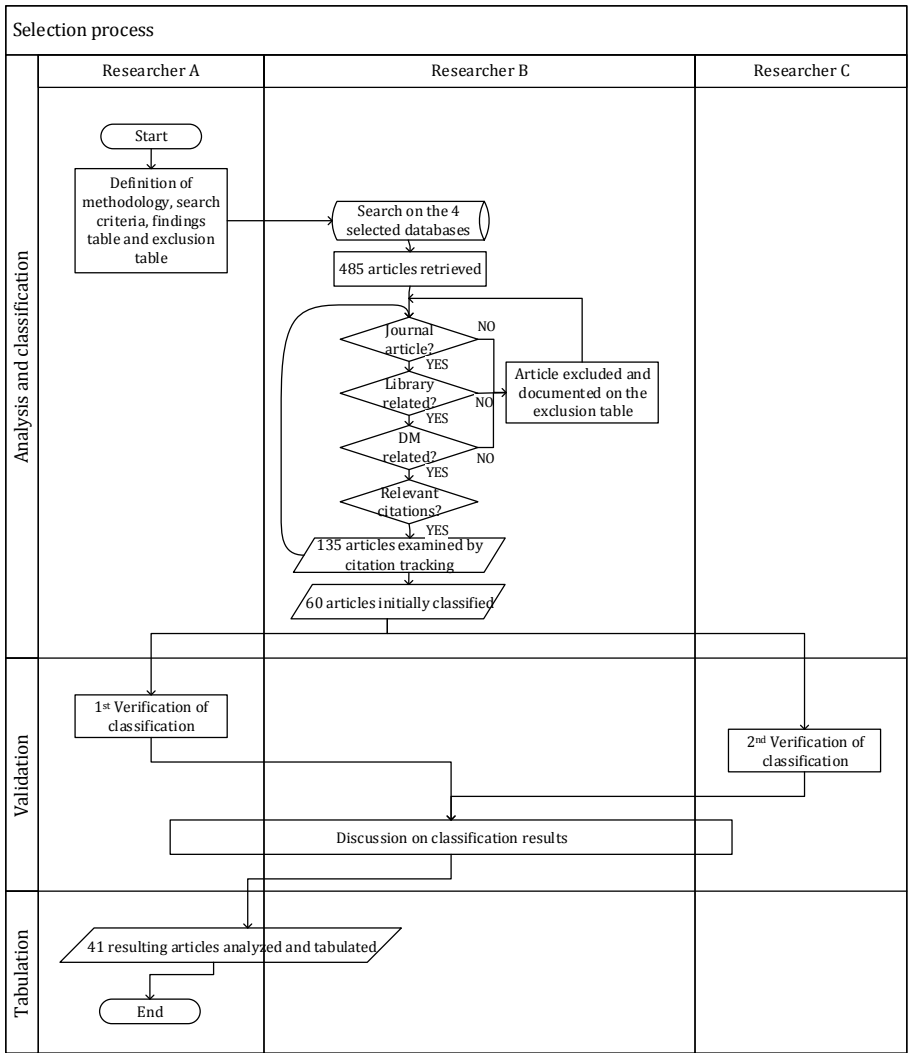


Figure 7.3: Selection process framework

7.4.1 Distribution of articles by the library holistic quadrants and data mining models

The distribution of articles classified by the proposed classification model is shown in Table 7.3. It is striking that a large part of the published case studies on the use of data mining in libraries are case studies of usage behavior analysis (24 out of 41 articles, 59%). Of these 24 articles, almost 38% of the articles (nine in total) are related to the analysis or characterization of data. For instance, log analysis is reported in four articles to analyze the information seeking behavior in regard to digital libraries and library websites. Specifically, Blecic et al. (1998) employ transaction log analysis of an OPAC and statistical tools to improve information retrieval. Nicholas et al. (2006)

Table 7.2: Distribution of articles according to the proposed classification model

Library holistic evaluation	Data mining functions	Data mining techniques	References
Service analysis	Association	Association rules	Decker and Höppner (2006)
	Classification	Logical analysis of data	Leonard et al. (2010), Zweibel and Lane (2012), Sewell (2013)
	Clustering	k-means algorithm	Tempelman-Kluit and Pearce (2014)
	Regression	Linear regression	Walters (2007), Weiner (2009), Emmons and Wilkinson (2011)
		Logistic regression	Yi (2009), Yi (2011), Yi (2012)
Quality evaluation		Regression analysis	Siriprasoetsin et al. (2011)
	Classification	Neural network	Decker and Hermelbracht (2006), Papavaslopoulos and Poulos (2012)
	Regression	Linear regression	Whitmire (2002)
		Regression analysis	Siriprasoetsin et al. (2011)
	Association	Association rules	Wu et al. (2004), Zhang and Wang (2013), Li (2014)
Collection analysis		Bibliometric analysis	Will (2006)
		Statistical analysis	Tosaka and Weng (2011)
	Classification	Decision/Classification tree	Kao et al. (2003), Nicholson (2003a), Wu (2003), King et al. (2007), Yang (2012)
		Log analysis	Nicholas et al. (2006)
		Logical analysis of data	Leonard et al. (2010), Zweibel and Lane (2012)
		Logistic regression	Nicholson (2003a)
		Memory-based reasoning	Nicholson (2003a)
		Neural network	Nicholson (2003a), Papavaslopoulos and Poulos (2012)
	Clustering	Decision/Classification tree	Koulouris and Kapidakis (2012)
		Pattern based clustering	Shreeves et al. (2003)
Regression		Linear regression	Walters (2007), Emmons and Wilkinson (2011)
		Logistic regression	Soria et al. (2014)

Library holistic evaluation	Data mining functions	Data mining techniques	References
Usage analysis	Association	Association rules	Pu and Yang (2003), Wu et al. (2004), Decker and Höppner (2006), Zhang and Wang (2013)
		Log analysis	Blecic et al. (1998)
		Statistical analysis	Blecic et al. (1998), Tosaka and Weng (2011)
	Classification	Decision/Classification tree	King et al. (2007)
		Log analysis	Nicholas et al. (2006), Shieh (2012), Ahmad et al. (2014)
		Logical analysis of data	Samson (2014)
		Neural network	Decker and Hermelbracht (2006)
	Clustering	Statistical analysis	Shieh (2012)
		Hierarchical cluster analysis	Bollen and Luce (2002), Hájek and Stejskal (2014)
		k-means algorithm	Bollen and Luce (2002), Hájek and Stejskal (2014), Tempelman-Kluit and Pearce (2014)
		Logical analysis of data	Finnell and Fontane (2010)
	Regression	Pattern based clustering	Papathodorou et al. (2003), Shreeves et al. (2003), Todorinova et al. (2011)
		Linear regression	Weiner (2009), Emmons and Wilkinson (2011), Fagan (2014)
		Logistic regression	Bracke (2004), Soria et al. (2014)

* Remark: Each article may have used more than one data mining technique and may have been implemented in more than one library holistic quadrant; thus, it may appear more than once

report a deep log investigation of the use and users of the Blackwell Synergy, a proprietary interdisciplinary digital library. Ahmad et al. (2014) utilize deep log analysis techniques to evaluate the user acceptance of e-book adoption. Shieh (2012) utilizes log analysis and statistical tools to evaluate the usability and findability of library websites. Statistics tools are employed in the total of three articles, which include two of the above-described studies: Bleicic et al. (1998), Shieh (2012), and the study presented by Tosaka and Weng (2011) using statistical tools to examine the effect of content-enriched records on library materials usage. Logical analysis of data is discussed in two articles: Finnell and Fontane (2010) employed these tools to investigate the feasibility of using reference questions as a tool in the construction of study guides, instructional outreach, and collection development. In a recent study, Samson (2014) analyzes the value of library resources to institutional teaching and research needs through the usage study of library e-resources.

Table 7.3: Distribution of articles by the library holistic quadrant and data mining models

Holistic evaluation quadrants	Number per holistic quadrant	Data mining functions	Amount per data mining function
Service analysis	12	Association	1
		Classification	3
		Clustering	1
		Regression	7
Quality evaluation	4	Classification	2
		Regression	2
Collection analysis	19	Association	5
		Classification	9
		Clustering	2
		Regression	3
Usage analysis	24	Association	6
		Classification	6
		Clustering	7
		Prediction	5

* Remark: Each article may have used more than one data mining technique, and may have been implemented in more than one library holistic quadrant

More sophisticated data mining techniques used in this quadrant include: association rules (four out of 24 articles), linear regression (three articles in total), k-means algorithm (3 articles), and pattern based clustering (three articles). Regarding association rules, Pu and Yang (2003) provide new basis for information organization and retrieval applications. Authors utilize circulation patterns of similar users to discover association classes scattered across different subject hierarchies. Wu et al. (2004) use circulation statistics and association rule discovery to support decision-making for material acquisitions. Specifically, association rules are employed to open up the relationship between pairs of material categories to predict the users' needs. Decker and Höppner (2006) apply an association rules-based approach to explore the use of customer intelligence to support strategic planning processes using data warehouse tools. Zhang and Wang (2013) report the implementation of association rules to mine transactional data generated in the process of library service. The aim of this study is to provide accurate service for readers based on a user behavior analysis. Linear regression models are utilized by several authors to demonstrate the library's value. For instance, Weiner (2009) utilizes multiple regression analysis to analyze the library contribution to the University reputation, while Emmons and Wilkinson (2011) apply a linear regression model to evaluate the impact of academic libraries on students' persistence. Fagan (2014) has recently used linear regression analyses to explore relationships among several variables thought to predict full-text article requests, such as reference transactions, library instruction, database searches, and ongoing expenditures. Concerning the implementation of k-means algorithms, Bollen and Luce (2002) and Hájek and Stejskal (2014) report the implementation of two types of cluster analysis: hierarchical cluster analysis and k-means clustering to analyze user retrieval patterns in digital libraries. Bollen and Luce (2002) analyze the retrieval habits of users in order to assess the impact of a library collection and to determine the

structure of a given user community. Hájek and Stejskal (2014) try to identify the user behavior of a typical consumer to support library management ensuring the provision of the appropriate level of library services. Tempelman-Kluit and Pearce (2014) utilize a k-means cluster to mine a Library 2.0 service. Chat reference data are analyzed to create hypothetical users (personas) that represent behaviors, goals and values of actual users. Eventually, Papatheodorou et al. (2003) use pattern based clustering to construct user communities sharing common interests and preferences. To do so, Z39.50 session log files are recorded and mined. Shreeves et al. (2003) identifies document clusters of potential interest, and provides visual displays of these clusters and document similarities. This study is part of a bigger project to examine the efficacy of using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to construct a search and discovery service focused on information resources in the domain of cultural heritage. Todorinova et al. (2011) examine the staffing patterns at the reference desk in order to give librarians greater flexibility as well as to allow better responding to the information-seeking needs of users.

Furthermore, 19 out of 41 articles (46%) deal with the application of data mining models in collection analysis, 12 articles (29%) with service analysis, and four articles (10%) with quality analysis, thus covering various aspects of library services and collection.

Articles covering the holistic quadrants of services, collection, and usage analysis apply all four data mining functions, whereas collection analyses do not employ cluster algorithms for their analyses. Collection and usage analyses are the two quadrants that have been the most explored together (nine articles). The majority of articles regarding quality analysis also cover the other three library aspects (three out of four articles). In fact, quality is the quadrant with the least-independent works, whereas usage behavior is the quadrant with the highest number of independent works (12 out of 24 articles).

Within the 24 articles of usage analysis, implementation of the data mining functions are almost equally distributed among them; that is, seven articles (29%) use clustering models to analyze the usage behavior of library collection, followed by association models and classification rules that are both discussed in six articles (25%) each, and five articles (21%) which use regression models. Regarding collection analysis, 47% (nine out of 19 articles) use classification models, and 26% (five articles) utilize association models. Figure 7.4 shows a visual representation of the classification of data mining applications based on the quadrants of the holistic evaluation matrix.

Table 7.4 shows the distribution of articles by data mining techniques. Among 14 data mining techniques, which have been applied in libraries, logistic regression is the most commonly used technique (six out of 41 articles), followed closely by association rules, decision/classification tree, linear regression, and logical analysis of data (five articles each). Among the top 10 data mining techniques, log analysis is described in four articles, and statistical analysis, k-means clustering and pattern based clustering are each described in three articles.

Although logistic regression is implemented in the majority of articles, linear regression is the technique with the highest frequency among library quadrants, and in turn the only technique implemented in all library holistic quadrants. Furthermore, regression models are the techniques that have been constantly implemented from the beginning. Collection and usage analyses both utilize 11 different types of algorithms (out of 14). Nevertheless, usage analysis is the quadrant that most data mining techniques have employed (28 techniques).

7.4.2 Distribution of articles by year of publication

The distribution of articles by year of publication and country of implementation is depicted in Table 7.5. One of the first reports on the use of data mining techniques in libraries is provided by Banerjee (1998). The author describes data mining functionalities, raises issues with their use, and discusses prerequisites for successfully utilizing data mining in libraries. Over the last six years (2009 to 2014), a positive increasing trend can be observed (see Figure 7.5), except for 2013 in which only two articles were published.

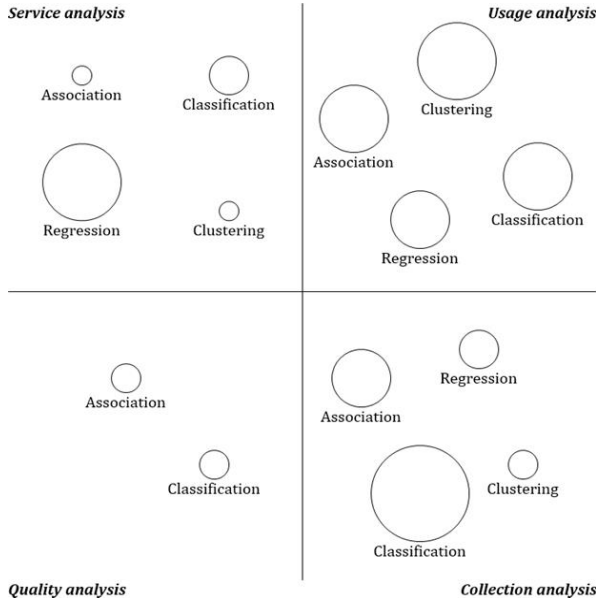


Figure 7.4: Classification of data mining applications based on the holistic evaluation matrix

Table 7.4: Distribution of articles by data mining techniques and library holistic quadrants

Data mining techniques	Service analysis	Quality analysis	Collection analysis	Usage analysis	Frequency	Number of articles	Percentage (%)
Logistic regression	4		2	2	8	6	15
Association rules	1		3	4	8	5	11
Decision/Classification tree			5	1	6	5	11
Logical analysis of data	3		2	2	7	5	11
Linear regression	4	1	2	4	11	5	11
Log analysis			1	4	5	4	9
K-means algorithm	1			3	4	3	7
Pattern based clustering			1	3	4	3	7
Statistical analysis			1	2	3	3	7
Hierarchical cluster analysis				2	2	2	4
Neutral network		2	1	1	4	2	4
Bibliometric analysis			1		1	1	2
Memory-based reasoning			1		1	1	2
Regression analysis	1	1			2	1	2
<i>Count: 14</i>	<i>14</i>	<i>4</i>	<i>20</i>	<i>28</i>	<i>Total:</i>	<i>41</i>	<i>100.00</i>

* Remark: Each article may have used more than one data mining technique and may have been implemented in more than one library holistic quadrant

Among the 16-year publication analysis, it is remarkable that in 2003 and 2006, the number of publications increased from an average of about two papers per year to six and four papers respectively, when compared to other years. Despite the fact that the first publication on the topic was on 1998, this trend reflects that the true first efforts of implementation of data mining functions in libraries were carried on from 2002. In 2003, three out of six articles report a case

study implementation in Taiwan and two articles in the USA. In Taiwan, Kao et al. (2003), Wu (2003), and Wu et al. (2004) led the implementation of data mining techniques in libraries by developing a knowledge management framework that utilizes data mining of circulation data to assess use of materials by particular academic departments in their subject areas. The techniques utilized in these studies are decision tree and association rules. In the USA, important to highlight is the study presented by Nicholson (2003a) that compares the effectiveness of four different data mining functions: logistic regression, memory-based reasoning, decision/classification tree and neural networks to discover Web-based scholarly research works. Moreover, in 2006, two out of four articles report a case study implementation in Germany, all implemented by Decker and colleagues (Decker & Hermelbracht, 2006; Decker & Höppner, 2006).

Table 7.5: Distribution of articles by year of publication and country of implementation

	Australia	China	Czech Republic	Germany	Greece	Netherlands	Taiwan	Thailand	UK	USA	Total
1998										1	1
1999											0
2000											0
2001											0
2002										2	2
2003					1		3			2	6
2004							1			1	2
2005											0
2006				2		1			1		4
2007	1									1	2
2008											0
2009										2	2
2010										3	3
2011								1		3	4
2012					2		2			2	6
2013		1								1	2
2014	1	1	1							4	7
Total	2	2	1	2	3	1	6	1	1	22	41

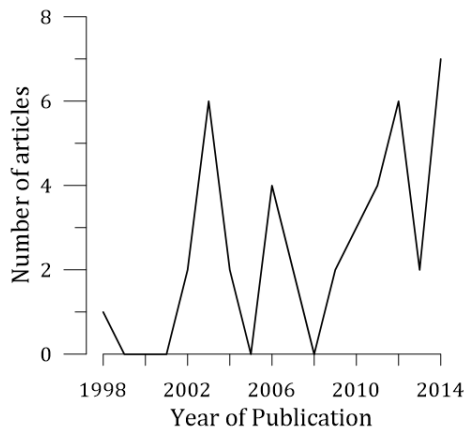


Figure 7.5: Evolution of number of articles per year

Data analysis and characterization are the most used techniques, except for the years 2002 – 2004, in which more formal data mining techniques are implemented. Quality is the only quadrant that has not employed these classical techniques as data mining tools. Important to note is that in the year 2014, articles have utilized the majority of data mining techniques, being 70% related to more formal procedures.

Usage analyses show an average of 1.4 publications per year throughout the period analyzed (see Figure 7.6). Strikingly, is also that in 2014, an increased interest in usage analyses is observed, since 6 out of 7 articles published in this year utilize several data mining functions, such as regression analysis, data characterization and cluster analysis, to analyze the usage of library e-resources, as well as to understand the information seeking behavior of users. Collection analyses are the second highest starting from 2003 onwards, especially in 2003 and 2012 with four out of seven articles and four out of eight articles published in those years respectively. Only isolated attempts to implement data mining functions in quality analyses can be observed in 2002, 2006, 2011, and 2012. Unfortunately, no studies have been reported on the use of data mining in quality analyses since 2012. Finally, articles focused on the use of data mining functions in service analyses emerged from 2006 onwards.

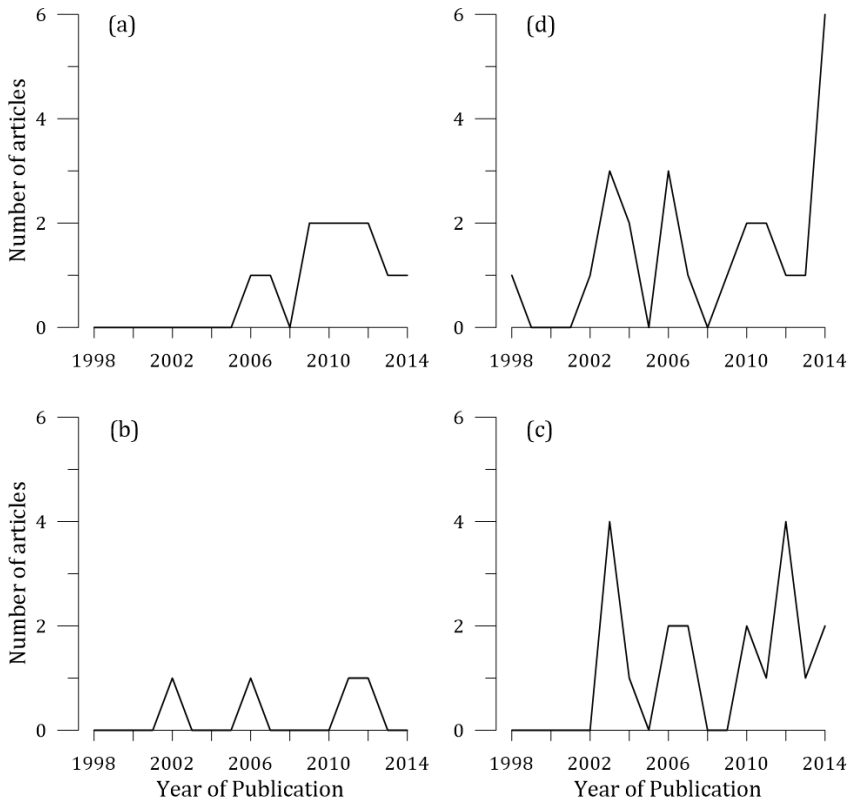


Figure 7.6: Chronological evolution of articles by library holistic evaluation. (a) Service analysis; (b) Quality analysis; (c) Collection analysis; (d) Usage analysis

7.4.3 Distribution of articles by country of implementation

Figure 7.7 shows the distribution of articles by country of implementation. Interesting is to highlight that the United States led by far among the list of countries that reported the implementation of data mining techniques (54% of case studies). Following in second is Taiwan, with a population of less than one tenth of the USA, with 15% of case studies implemented (six in total), and the third is Greece with 7% of case studies implemented (three out of 41). It is worth noting that four out of four articles covering both service and collection analyses are implemented in US academic libraries.

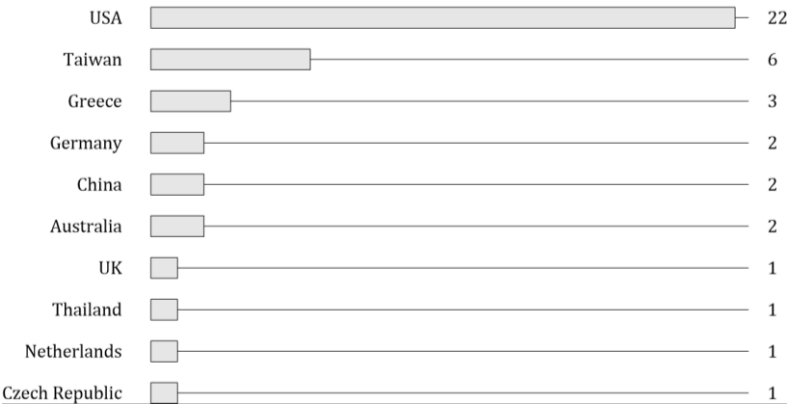


Figure 7.7: Distribution of articles by country of implementation

7.4.4 Distribution of articles by journal in which the articles were published

Figure 7.8 shows the top six journals, which contain the highest number of research articles. Articles related to application of data mining techniques in libraries are distributed across 27 journals. These findings indicate that scientific contributions in this research area are scattered across a high range of journals (average of 1.52 articles per journal), particularly related to computer science, and information and library management. The top six journals which the highest number of research articles, contain almost 50% (20 out of 41 articles) of the total number of articles published.

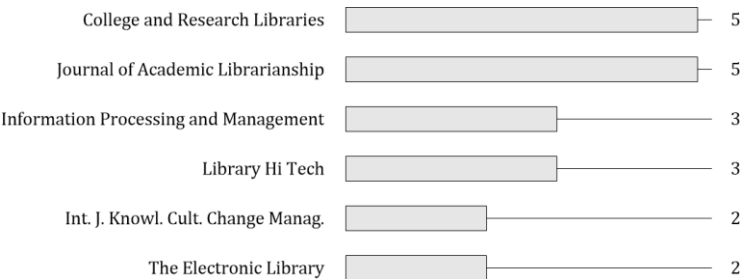


Figure 7.8: Distribution of articles in the top 6 journals

Of these, “College and Research Libraries” the official scholarly research journal of the Association of College & Research Libraries, and the “Journal of Academic Librarianship” both focused on problems and issues relevant to college and university libraries, each containing over 12% (five out of 41 articles) of the total number of articles published, followed by the “Information Processing and Management” and “Library Hi Tech” with three articles each, and the “International Journal of Knowledge, Culture and Change Management” and “The Electronic Library” journal with two articles each. All are related to libraries except for the third and fifth ranked journal; the third journal is IT related, while the fifth is Management related.

The four databases (LISA, LISTA, WoS, and Scopus) were rechecked to determine where the articles were indexed. Scopus is the multidisciplinary database that provides the best coverage of journal articles identified in this study with 39 articles in total, followed by WoS with 34 articles found. LISA and LISTA index 26 and 27 articles out of 41 respectively. Evidently, the combination of the online databases allowed for the gathering of the 41 analyzed articles.

7.4.5 Distribution of articles by library type analyzed

Figure 7.9 shows the distribution of articles by type of library analyzed: physical, digital, or both, through data mining techniques. Of the 41 articles, 41% (17 articles) are related to the use of data mining techniques in both digital and physical libraries, and 37% (15 articles) are related to the use of data mining techniques in digital libraries. This result is not unexpected, and confirms the natural transition and evolving trend of shifting the focus from physical to digital collection and services. Important to note is that not all the articles clearly specified the type of library analyzed, therefore, a certain degree of subjectivity can be present. In addition, three articles reported by Yi (2009, 2011; 2012) are not specifically focused on a specific library type, since all examine how academic library directors plan and manage change in information technology and the factors influencing the planning and management approaches used.

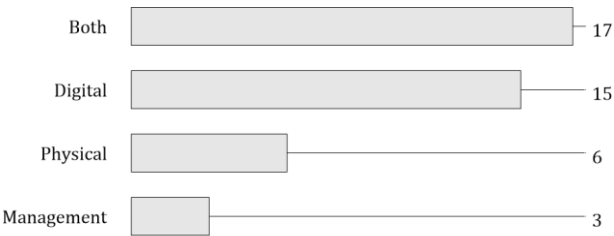


Figure 7.9: Distribution of articles by type of library analyzed

7.5 Limitations

The methodology that is employed in the literature review and classification of data mining techniques in libraries has some limitations. The first is that the study analyzes articles extracted based on specific keywords such as “data mining”, “case studies”, “academic librar*”, “university librar*” and “bibliomining”. Articles which mentioned the application of data mining techniques in academic libraries without these keywords may have been omitted during the retrieval process. The second is that findings are based on data collected only from academic journals, so other materials which may contain more case studies on data mining applications might have been excluded. The third is the limited number of databases used (two multidisciplinary and two LIS oriented), and from these, only the journals in the particular databases that were searched were included. However, although this limitation could mean that the review is not exhaustive, the authors believe that it is comprehensive by providing reasonable insights into the work being

accomplished in the area, and also because the databases selected are the most important journal databases in their corresponding domain. The fourth possible limitation is that the articles' classification process was subjective; nevertheless, research triangulation allowed for a reduction of this risk. Finally, the last limitation is that the study includes only English publications; by so doing, this restriction could jeopardize the analysis since surely more research regarding the application of data mining techniques in libraries is being discussed and published in other languages.

7.6 Conclusion and research implications

The application of data mining techniques in libraries is an emerging trend that has captured the attention of practitioners and academics in order to understand patterns of behavior of library users and staff, and patterns of information on resource usage throughout the library. The aim of this literature review has been to facilitate and ease the interested reader or practitioner's introduction to the use of data mining in libraries. To do so, the article presents a comprehensive literature review on the implementation of data mining techniques in libraries with a special focus on the case studies published over the period 1998-2014. Forty-one papers were identified, analyzed and classified along the four quadrants of the holistic evaluation matrix that analyze services, quality, collection, and usage behavior in libraries; and the main data mining functions, which are, clustering, association, classification, and regression. Although this literature review cannot claim to be exhaustive, it does highlight important implications, as well as insights into the state-of-the-art. For instance:

- Nicholson's idea of coining the term "bibliomining" to refer to the use of data mining in libraries (Nicholson & Stanton, 2003) was an important contribution in classifying this rapidly emerging topic. Data mining in libraries can be defined as the core of a larger process dubbed as bibliomining. Thus, the use of data mining to examine library data records might be aptly termed bibliomining. Unfortunately, in practice, only few researchers have used the "bibliomining" term in their publications (3 articles), and consequently, it cannot be considered as a standard word.
- According to past publication rates and the increasing interest in the use of data mining tools in libraries, practical research will increase significantly in this area in the future, and consequently, a significant increase in research and published literature is expected.
- Among the reviewed articles, all of them use one or two data mining techniques to analyze only one or two library holistic quadrants, and just one case study in the literature, by Emmons and Wilkinson (2011), has reported a case study covering the analysis of three library holistic quadrants: process, collection, and usage. None of the case studies cover the four library evaluation quadrants. Knowing that a combination of data mining models and library evaluation quadrants is often required to solve, support, or forecast the effects of library strategies, library directors should include more data mining functions to support their holistic-based decision-making.
- The majority of the reviewed articles relate to usage analysis. Of these, about 38% (nine out of 24 articles) discuss data analysis techniques such as logical analysis of data and analysis of logs and statistics, 33% (eight articles) use cluster analysis, 29% (seven articles) utilize supervised learning tools and the remaining 17% (four articles) analyze dependences through association rules. The main library aspects covered through these studies are the interaction of library users with the system, the usability of library websites, and the users' categorization based on the usage interaction with the system and collection.
- Only a few articles of the 41 reviewed are related to quality analysis (four articles in total). The small number of research was somewhat surprising given that libraries have a long history on collecting statistics to answer users' queries, and thus monitor service quality (Horn & Owen, 2009); however, this topic is scarcely covered in the LIS literature since only a limited number of articles have reported the usage of sophisticated quality analysis as shown in this study. Further research needs to be conducted in this area, especially in quality control or in considering quality as an important factor when implementing data mining functions in other library aspects.

- Findings indicate that service analysis is slowly emerging as a possible new domain from this research. This library aspect is a crucial element for successful decision-making, especially due to increasingly difficult times, characterized by budget constraints and dynamic services. More than ever, libraries need to demonstrate that their processes and inputs such as facilities, expenditures, and staffing are considered relevant and worthwhile in their outputs through data on services and people served. Therefore, more research is highly recommended on the use of data mining techniques in the analysis of service performance in both digital and physical environments.
- During the research a common theme that has emerged was the appropriated definition of "data mining". Actually, the concept is difficult to explain and several authors opine that the term is a misnomer and a buzzword (Han et al., 2011; J. Wu, 2012). In this study, 10 out of 11 articles (almost 25% of 41 articles) implementing data analysis and characterization approaches, such as logical analysis of data, and analyses of statistics, logs and bibliometrics, include as part of their topic terms (title, keywords or even descriptors) the words "data mining". The reasoning behind the inclusion of these articles, which can be argued to be not data mining techniques, is to highlight the overuse of the data mining concept in the LIS literature. Therefore, to benefit from the advantages of data mining, it is recommended that further studies be conducted utilizing more enhanced techniques that have not been documented previously such as super vector machines and ensemble methods.
- Among the 41 case studies reviewed for this article, 14 articles utilize classification models and 11 use regression techniques to assist in library decision-making. Laggards are the implementation of unsupervised algorithms such as association and clustering models (eight out of 41 articles each). Knowing that unsupervised learning allows finding hidden structure in unlabeled data, as well as allows spotting salient correlations and connections between data points that are not evident for humans, the implementation of further association and clustering algorithms is highly recommended.
- Association rules and decision/classification trees rank after logistic regression in popularity of application in libraries. The logic of both techniques can be followed more easily by librarians and information specialists. Therefore, the two techniques are highly recommended for non-experts in data mining techniques.
- The top six journals which contain almost 50% of the total number of articles published on the application of data mining techniques in academic libraries are: College and Research Libraries, Journal of Academic Librarianship, Information Processing and Management, Library Hi Tech, International Journal of Knowledge, Culture and Change Management, and The Electronic Library. Scopus is the multidisciplinary database that indexes almost all articles identified in this study.
- The majority of research articles have been implemented in the United States.
- Findings indicate that the attention of implementing data mining techniques in library management literature has mainly been directed towards digital collection and e-services (37%), and less towards physical collection (15%). This is not surprising as digital libraries are becoming more and more prevalent worldwide.

References

- ACRL Research Planning and Review Committee. (2014). Top trends in academic libraries: A review of the trends and issues affecting academic libraries in higher education. *College & Research Libraries News*, 75(6), 294–302.
- Ahmad, P., Brogan, M., & Johnstone, M. N. (2014). The e-book power user in academic and research libraries: Deep log analysis and user customisation. *Australian Academic & Research Libraries*, 45(1), 35–47. <http://doi.org/10.1080/00048623.2014.885374>
- Association of College and Research Libraries. (2010). Value of academic libraries: A comprehensive research review and report. Chicago, USA: Association of College and Research Libraries.
- Banerjee, K. (1998). Is data mining right for your library? *Computers in Libraries*, 18(10), 28–31.
- Blecic, D. D., Bangalore, N. S., Dorsch, J. L., Henderson, C. L., Koenig, M. H., & Weller, A. C. (1998). Using transaction log analysis to improve OPAC retrieval results. *College & Research Libraries*, 59(1), 39–50. <http://doi.org/10.5860/crl.59.1.39>

- Bollen, J., & Luce, R. (2002). Evaluation of digital library impact and user communities by analysis of usage patterns. *D-Lib Magazine*, 8(6). <http://doi.org/10.1045/june2002-bollen>
- Bracke, P. J. (2004). Web usage mining at an academic health sciences library: an exploratory study. *Journal of the Medical Library Association*, 92(4), 421–428.
- Chang, C.-C., & Chen, R.-S. (2006). Using data mining technology to solve classification problems: A case study of campus digital library. *Electronic Library*, 24(3), 307–321. <http://doi.org/10.1108/02640470610671178>
- Chen, S. Y., & Liu, X. (2004). The contribution of data mining to information science. *Journal of Information Science*, 30(6), 550–558. <http://doi.org/10.1177/0165551504047928>
- Decker, R., & Hermelbracht, A. (2006). Planning and evaluation of new academic library services by means of Web-based conjoint analysis. *The Journal of Academic Librarianship*, 32(6), 558–572. <http://doi.org/10.1016/j.acalib.2006.06.016>
- Decker, R., & Höppner, M. (2006). Strategic planning and customer intelligence in academic libraries. *Library Hi Tech*, 24(4), 504–514. <http://doi.org/10.1108/07378830610715374>
- Emmons, M., & Wilkinson, F. C. (2011). The Academic Library Impact on Student Persistence. *College & Research Libraries*, crl-74r1.
- Fagan, J. C. (2014). The Effects of Reference, Instruction, Database Searches, and Ongoing Expenditures on Full-text Article Requests: An Exploratory Analysis. *The Journal of Academic Librarianship*, 40(3–4), 264–274. <http://doi.org/10.1016/j.acalib.2014.04.002>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Commun. ACM*, 39(11), 27–34. <http://doi.org/10.1145/240455.240464>
- Finnell, J., & Fontane, W. (2010). Reference Question Data Mining: A Systematic Approach to Library Outreach. *Reference & User Services Quarterly*, 49(3), 278–286.
- Girija, N., & Srivatsa, S. K. (2006). A research study: Using Data Mining in knowledge base business strategies. *Information Technology Journal*, 5(3), 590–600. <http://doi.org/10.3923/itj.2006.590.600>
- Gonzalez, R., Llopis, J., & Gasco, J. (2013). Information systems offshore outsourcing: managerial conclusions from academic research. *International Entrepreneurship and Management Journal*, 9(2), 229–259. <http://doi.org/10.1007/s11365-013-0250-y>
- Hájek, P., & Stejskal, J. (2014). Library user behavior analysis - Use in economics and management. *Wseas Transactions on Business and Economics*, 11, 107–116.
- Hand, D. J. (1998). Data Mining: Statistics and more? *The American Statistician*, 52(2), 112–118. <http://doi.org/10.1080/00031305.1998.10480549>
- Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining*. MIT Press.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and techniques* (3rd ed.). Elsevier.
- Horn, A., & Owen, S. (2009). Mind the gap 2014 : research to inform the next five years of library development. In *Innovate, collaborate : conference proceedings EDUCAUSE Australasia 2009* (pp. 1–12). Perth, Western Australia. Retrieved from <http://dro.deakin.edu.au/view/DU:30016487>
- Hui, S. C., & Jha, G. (2000). Data mining for customer service support. *Information & Management*, 38(1), 1–13. [http://doi.org/10.1016/S0378-7206\(00\)00051-3](http://doi.org/10.1016/S0378-7206(00)00051-3)
- Kao, S.-C., Chang, H.-C., & Lin, C.-H. (2003). Decision support for the academic library acquisition budget allocation via circulation database mining. *Information Processing & Management*, 39(1), 133–147. [http://doi.org/10.1016/S0306-4573\(02\)00019-5](http://doi.org/10.1016/S0306-4573(02)00019-5)
- King, J. D., Li, Y., Tao, X., & Nayak, R. (2007). Mining world knowledge for analysis of search engine content. *Web Intelligence and Agent Systems*, 5(3), 233–253.
- Kononenko, I., & Kukar, M. (2007). *Machine Learning and Data Mining*. Elsevier.
- Koulouris, A., & Kapidakis, S. (2012). Policy route map for academic libraries' digital content. *Journal of Librarianship and Information Science*, 44(3), 163–173. <http://doi.org/10.1177/0961000612444299>
- Leonard, M. F., Haas, S. C., & Kissing, V. N. (2010). Metrics and science monograph collections at the Marston science library, University of Florida. *Issues in Science and Technology Librarianship*, 62. Retrieved from <http://dialnet.unirioja.es/servlet/articulo?codigo=3706724>
- Li, X. (2014). An Algorithm for Mining Frequent Itemsets from Library Big Data. *Journal of Software*, 9(9), 2361–2365. <http://doi.org/10.4304/jsw.9.9.2361-2365>

- Lunfeng, G., Huan, L., & Li, Z. (2012). The application of association rules of data mining in book-lending service. In 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD) (pp. 761–764). <http://doi.org/10.1109/FSKD.2012.6233921>
- Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2, Part 2), 2592–2602. <http://doi.org/10.1016/j.eswa.2008.02.021>
- Nicholas, D., Huntington, P., Monopoli, M., & Watkinson, A. (2006). Engaging with scholarly digital libraries (publisher platforms): The extent to which “added-value” functions are used. *Information Processing & Management*, 42(3), 826–842. <http://doi.org/10.1016/j.ipm.2005.03.019>
- Nicholson, S. (2003a). Bibliomining for automated collection development in a digital library setting: Using data mining to discover Web-based scholarly research works. *Journal of the American Society for Information Science and Technology*, 54(12), 1081–1090. <http://doi.org/10.1002/asi.10313>
- Nicholson, S. (2003b). The bibliomining process: Data Warehousing and Data Mining for library decision-making. *Information Technology and Libraries*, 22(4), 146–151.
- Nicholson, S. (2004). A conceptual framework for the holistic measurement and cumulative evaluation of library services. *Journal of Documentation*, 60(2), 164–182. <http://doi.org/10.1108/00220410410522043>
- Nicholson, S. (2006a). Approaching librarianship from the data: Using bibliomining for evidence-based librarianship. *Library Hi Tech*, 24(3), 369–375. <http://doi.org/10.1108/07378830610692136>
- Nicholson, S. (2006b). The basis for bibliomining: Frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services. *Information Processing & Management*, 42(3), 785–804. <http://doi.org/10.1016/j.ipm.2005.05.008>
- Nicholson, S., & Stanton, J. (2006). Bibliomining for library decision-making. In *Encyclopedia of Data Warehousing and Mining* (Second Edition, pp. 100–105). Retrieved from <http://www.igi-global.com/chapter/encyclopedia-data-warehousing-mining/10591>
- Nicholson, S., & Stanton, J. M. (2003). Gaining strategic advantage through Bibliomining: Data Mining for management decisions in corporate, special, digital, and traditional libraries. In *Organizational data mining: Leveraging enterprise data resources for optimal performance*. Hershey, PA: Idea Group Publishing. Retrieved from http://arizona.openrepository.com/arizona/bitstream/10150/106383/1/Nicholson_3.pdf
- Papatheodorou, C., Kapidakis, S., Sfakakis, M., & Vassiliou, A. (2003). Mining user communities in digital libraries. *Information Technology and Libraries*, 22(4), 152–157.
- Papavaslopoulos, S., & Poulos, M. (2012). Neural network design and evaluation for classifying library indicators using personal opinion of expert. *Library Management*, 33(4/5), 261–271. <http://doi.org/10.1108/01435121211242308>
- Prakash, K., Chand, P., & Gohel, U. (2004). Application of Data Mining in library and information services (pp. 168–177). Presented at the 2nd Convention PLANNER, Manipur Uni., Imphal: INFLIBNET Centre, Ahmedabad. Retrieved from <http://shodhganga.inflibnet.ac.in/dxml/handle/1944/435>
- Pu, H., & Yang, C. (2003). Enriching user-oriented class associations for library classification schemes. *The Electronic Library*, 21(2), 130–141. <http://doi.org/10.1108/02640470310470507>
- Robin R. Sewell. (2013). Who is following us? Data mining a library's Twitter followers. *Library Hi Tech*, 31(1), 160–170. <http://doi.org/10.1108/07378831311303994>
- Samson, S. (2014). Usage of e-resources: Virtual value of demographics. *The Journal of Academic Librarianship*, 40(6), 620–625. <http://doi.org/10.1016/j.acalib.2014.10.005>
- Shieh, J.-C. (2010). The integration system for librarians' bibliomining. *Electronic Library*, The, 28(5), 709–721. <http://doi.org/10.1108/02640471011081988>
- Shieh, J.-C. (2012). From website log to findability. *Electronic Library*, The, 30(5), 707–720. <http://doi.org/10.1108/02640471211275747>
- Shreeves, S. L., Kaczmarek, J. S., & Cole, T. W. (2003). Harvesting cultural heritage metadata using the OAI Protocol. *Library Hi Tech*, 21(2), 159–169. <http://doi.org/10.1108/07378830310479802>

- Siguenza-Guzman, L., Van Den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2015). A holistic approach to supporting academic libraries in resource allocation processes. *The Library Quarterly: Information, Community, Policy*, 85 (3).
- Siriprasoetsin, P., Tuamsuk, K., & Vongprasert, C. (2011). Factors affecting customer relationship management practices in Thai academic libraries. *The International Information & Library Review*, 43(4), 221–229. <http://doi.org/10.1016/j.iilr.2011.10.008>
- Soria, K. M., Fransen, J., & Nackerud, S. (2014). Stacks, Serials, Search Engines, and Students' Success: First-Year Undergraduate Students' Library Use, Academic Achievement, and Retention. *The Journal of Academic Librarianship*, 40(1), 84–91. <http://doi.org/10.1016/j.acalib.2013.12.002>
- Sumathi, S., & Sivanandam, S. N. (2006). *Introduction to Data Mining and its applications*. Springer.
- Tempelman-Kluit, N., & Pearce, A. (2014). Invoking the User from Data to Design. *College & Research Libraries*, 75(5), 616–640. <http://doi.org/10.5860/crl.75.5.616>
- Todorinova, L., Huse, A., Lewis, B., & Torrence, M. (2011). Making Decisions: Using Electronic Data Collection to Re-Envision Reference Services at the USF Tampa Libraries. *Public Services Quarterly*, 7(1-2), 34–48. <http://doi.org/10.1080/15228959.2011.572780>
- Tosaka, Y., & Weng, C. (2011). Reexamining Content-Enriched Access: Its Effect on Usage and Discovery. *College & Research Libraries*, 72(5), 412–427. <http://doi.org/10.5860/crl-137>
- Walters, W. H. (2007). A Regression-based Approach to Library Fund Allocation. *Library Resources & Technical Services*, 51(4), 263–278.
- Weiner, S. (2009). The Contribution of the Library to the Reputation of a University. *The Journal of Academic Librarianship*, 35(1), 3–13. <http://doi.org/10.1016/j.acalib.2008.10.003>
- Whitmire, E. (2002). Academic library performance measures and undergraduates' library use and educational outcomes. *Library & Information Science Research*, 24(2), 107–128. [http://doi.org/10.1016/S0740-8188\(02\)00108-1](http://doi.org/10.1016/S0740-8188(02)00108-1)
- Will, N. (2006). Data-mining: Improvement of university library services. *Technological Forecasting and Social Change*, 73(8), 1045–1050. <http://doi.org/10.1016/j.techfore.2006.05.006>
- Wright, S., & White, L. S. (2007). Library Assessment: SPEC Kit 303. *Association of Research Libraries*, 14.
- Wu, C.-H. (2003). Data mining applied to material acquisition budget allocation for libraries: design and development. *Expert Systems with Applications*, 25(3), 401–411. [http://doi.org/10.1016/S0957-4174\(03\)00065-4](http://doi.org/10.1016/S0957-4174(03)00065-4)
- Wu, C.-H., Lee, T.-Z., & Kao, S.-C. (2004). Knowledge discovery applied to material acquisitions for libraries. *Information Processing & Management*, 40(4), 709–725. <http://doi.org/10.1016/j.ipm.2003.08.010>
- Wu, J. (2012). *Advances in K-means Clustering: A Data Mining Thinking*. Springer Science & Business Media.
- Yang, S.-T. (2012). An active recommendation approach to improve book-acquisition process. *International Journal of Electronic Business Management*, 10(2), 163–73.
- Yi, Z. (2009). The management of change in information technology: Approaches of academic library directors in the United States. *International Journal of Knowledge, Culture and Change Management*, 9(11), 109–130.
- Yi, Z. (2011). Planning change in the information age: Approaches of academic library directors in the United States. *International Journal of Knowledge, Culture and Change Management*, 10(12), 155–176.
- Yi, Z. (2012). Conducting meetings in the change process: Approaches of academic library directors in the United States. *Library Management*, 33(1/2), 22–35. <http://doi.org/10.1108/01435121211203293>
- Zhang, Q. S., & Wang, X. Y. (2013). Research of Personalized Information Service Based on Association Rules. *Advanced Materials Research*, 760-762, 1800–1803. <http://doi.org/10.4028/www.scientific.net/AMR.760-762.1800>
- Zweibel, S., & Lane, Z. B. (2012). Probing the effects of policy changes by evaluating circulation activity data at Columbia University Libraries. *The Serials Librarian*, 63(1), 17–27. <http://doi.org/10.1080/0361526X.2012.687850>

Chapter 8: **An experimental use of**

Bibliomining

Siguenza-Guzman, L., Auquilla, A., Saquicela, V., Vandewalle, J., & Cattrysse, D. An experimental use of Bibliomining for Library Decision-Making. [Article in preparation]

This chapter reports an experimental use of data mining techniques for library decision-making based on data warehousing. First, the theoretical background and related work is briefly presented. Next, the chapter describes the selection of methodology for the construction of the DW, as well as the architecture and implementation framework. In this section, the three-layer data warehouse architecture described in Chapter 6 is used to integrate, process, and store the collected data. These stored data are then queried and analyzed utilizing three data mining techniques: regression, clustering and classification. This chapter finalizes by summarizing lessons learned and identifying future challenges and directions.

Abstract

This article reports an experimental use of bibliomining techniques to support library decision-making in an academic library. To do so, a three-layer data warehouse architecture is used to integrate, process, and store the collected data. Firstly, a holistic approach is used for data collection. This holistic approach incorporates key elements that may influence library decision-making, including service analysis, quality estimation, information relevance, and usage analysis. Secondly, the study uses the Hefesto methodology to implement a data warehouse, consisting of the following steps: requirement analysis, data source analysis, logical model definition, and data integration. Eventually, the stored data are analyzed through three bibliomining techniques: regression, clustering and classification. Specifically, regression techniques are used to predict future investments in service and collection development; clustering techniques are used to find clusters of users that share common interests and similar profiles, but belong to different units; and classification techniques are used to demonstrate library's value by analyzing possible correlations of library usage and academic performance.

Contributions of the first author

The first author's contributions are: the theoretical background, related work analysis, methodology for data warehouse implementation, architecture of the iDSS, data analysis, interpretation of results and conclusions.

8.1 Introduction

In this rapidly evolving, technologically-driven information environment, traditional library management systems are no longer sufficient to manage neither the new dynamic library services nor the large amounts of available information (Siguenza-Guzman, Van Den Abbeele, Vandewalle, Verhaaren, & Cattrysse, 2015; Xu & Li, 2013). Traditional library management systems are suitable to meet the automation needs of academic libraries in terms of operational activities and traditional physical services. Yet, to promote innovation and knowledge creation for strategic decision-making, traditional systems provide very limited functions. In this sense, knowledge management (KM) has become a powerful tool for libraries to expand their role to areas where they had little impact in the past such as financial decisions and strategic decision-making (Hobohm, 2004; Townley, 2001).

Knowledge-based decision support systems (DSS) provide important information for library decision-making and performance improvement (Lai, Wang, Huang, & Kao, 2011). Several DSS for libraries have been documented in literature; however most of them mainly focus on specific areas such as distribution of money for physical or digital collections, assessment of the performance of library collection or the analysis of user behavior. Little is known about integrating all these different aspects and incorporating others such as human resources, technological infrastructure, services, or library usage. Even less is known about embracing different perspectives from heterogeneous stakeholders such as academic staff, managers, librarians, and general users (Zhang, 2010). The purpose of this study is to present an experimental use of bibliomining techniques for library decision making under the context of an integrated decision support system (iDSS) based on a data warehouse (DW). The article begins by presenting the theoretical background and related work. Next, the article briefly describes the selection of methodology for the construction of the DW, as well as the architecture and implementation framework. The data stored in the DW are then queried and analyzed utilizing two data mining techniques: regression and clustering. The article finalizes by summarizing the lessons learned and identifying future challenges and directions.

8.2 Theoretical Background

In this section, the theoretical background of our study is developed through the literature review of the holistic data collection approach, DW and bibliomining.

8.2.1 Holistic data collection

An iDSS for library management represents a powerful tool for planning and controlling; unfortunately its implementation is very complex in practice. In fact, the main challenges to implement an iDSS are the number of data sources, formats and volume of data to be consulted, as well as the lack of standardized data collection structures (Siguenza-Guzman, Van Den Abbeele, et al., 2015). Regarding data collection, Scott Nicholson (2004) recommends a theoretical analysis framework that supports libraries in gaining a more thorough and holistic understanding of their users and services. The author proposes a two-dimension matrix that evaluates libraries based on their library system and collection, from an internal and external perspective. Due to the ease of understanding, completeness as well as applicability to both physical and digital resources, Siguenza Guzman, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., and Cattrysse, D. (2015) adopted the framework as the basis to propose a structure for data collection, an architecture for data storage and an integrated set of tools that assists library managers to make decisions through a holistic perspective¹⁵. An overview of holistic evaluation framework is shown in Figure 8.1.

¹⁵ More details on the architecture proposed and an integrated set of tools can be found in Siguenza-Guzman et al. (2015)

	<i>Library System</i>	<i>Library Collection</i>
<i>Internal Perspective (Library)</i>	1) Service Analysis Processes, cost, Time, Resources	4) Usage analysis Transaction log analysis, Deep log analysis
<i>External Perspective (Users)</i>	2) Quality Analysis Statistics gathering, Suggestion boxes, Usability testing, Satisfaction surveys	3) Collection Analysis Citation patterns, Publishing patterns, Journals downloaded, and Journal's impact factor

Figure 8.1: Methodologies proposed for the economic evaluation of libraries through a holistic approach (Siguenza-Guzman, Van Den Abbeele, et al., 2015)

The *first quadrant* analyzes the internal perspective of the library system that is to analyze the library performance, costs incurred and resources consumed by library services. For that purpose, three costing systems are analyzed: traditional, activity-based and time-driven activity-based, recommending the latter. The *second quadrant* evaluates the external perspective of the library system; quality of the processes and services offered by libraries are judged by the users. In this sense, the use of at least one of the following methods is recommended: statistics gathering, suggestion boxes, Web usability testing, user interface usability, and satisfaction surveys. The *third quadrant* analyzes the external perspective of the library collection, evaluating the impact of the current library collection on its users. The authors propose to combine the following three methods: citation analysis, vendor-supplied statistics and citation databases. Eventually, the *fourth quadrant* evaluates the internal perspective of the library collection, where the usage patterns followed to manipulate the library collection are analyzed. It is suggested to use log analysis methods such as transaction log analysis and deep log analysis.

The resulting framework requires retrieving and integrating information from various separate sources in order to be used in an adequate decision-making process. An example of preliminary results implementing this proposed holistic approach is presented in a case study by Siguenza Guzman, L., Holans, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., and Cattrysse, D. (2013). The authors prove the practical validity of the proposed holistic approach; and highlight the need of a system to integrate the collected information.

8.2.2 Data Warehouse

William H. Inmon, acknowledged as the father of "Data Warehouse", defines it as a "subject-oriented, integrated, nonvolatile, and time-variant collection of data in support of management's decisions" (Inmon, 2005, p. 29). According to Ralph Kimball (2006), the other preeminent figure in data warehousing, a DW is a repository of integrated information from distributed, autonomous, and possibly heterogeneous sources, specifically structured for analysis and consultation. A DW is a read-only analytical database that integrates all information from various operational data sources, whose purpose is to generate reports and analyze data in order to support the strategic decision-making process in an enterprise.

Data stored in the DW are snapshots resulting from data transformations, quality control checks and integration of operational data. The major benefit of a DW is the possibility to have interactive and immediate access to strategic information of an enterprise. Users with a managerial role in the organization make their own inquiries and crossing data, using specialized tools with graphical interfaces such as data mining and on-line analytical processing (OLAP) tools. The operations are not focused on recording transactions, as in the operational databases, but involve complex queries of joining, filtering, grouping, and aggregating large amounts of data (Wrembel & Koncilia, 2007).

Due to the special characteristics of a DW, the design strategies used for building and managing operational databases are generally not applicable to the design of DW (Inmon, 2005; Kimball, 2006). The design of a DW is an inherently complex, costly and time-consuming task. Building a DW, as shown in Figure 8.2, involves extracting data from different data sources, in which many problems of inconsistency need to be dealt with (Ying Wah, Hooi Peng, & Sue Hok, 2007). Data extraction, cleansing and storage through ETL (Extract, Transform, Load) processes is also complex and time-consuming, because this process needs to combine all the different data sources and converts them into a uniform format, excluding possible inconsistencies, redundancies, and incompatibilities (Nicholson, 2003). Therefore, to ensure the end results meet the needs, successful implementation of a DW requires, among other things, a significant investment of effort, from many of those who will be its users, to identify the business objectives, design the DW architecture and implement the DW system (Curtis & Joshi, 2011; Radulescu, 2003). However, as Scott Nicholson and Jeffrey Stanton (2009) remark, only by combining and linking different data sources, library managers can uncover hidden strategic information that can help to understand processes and services for library decision-making.

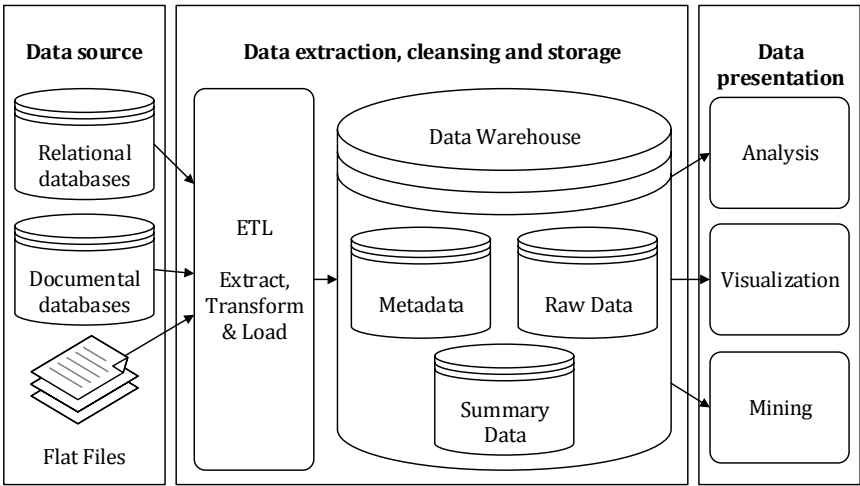


Figure 8.2: Architecture of a DW (Based on Inmon (2005))

8.2.3 Bibliomining

Bibliomining, or data mining for libraries, was first defined by Nicholson and Stanton (2003) as the combination of a DW, data mining and bibliometrics¹⁶ used to track patterns of behavior-based artifacts from library systems. Bibliomining is an important tool to discover unknown patterns and useful library information in historical data to support decision-making (Kao, Chang, & Lin, 2003). However, to provide a complete report of the library system, bibliomining needs to be used cyclically, based on a combination of data mining models and other methods of measurement and evaluation (Nicholson, 2003). Mining techniques in libraries, among others, generally include the following (Siguenza-Guzman, Saquicela, Avila, Vandewalle, & Cattrysse, 2015):

- a) Regression is a type of statistical technique that maps a data item into a real-valued prediction variable. This technique is intended to create a panorama of a phenomenon studied. Uses of regression techniques include forecasting and testing hypotheses about relationships between variables.

¹⁶ Bibliometrics is statistical analysis used to provide quantitative analysis of academic literature.

- b) Clustering is the task of uncovering unanticipated trends by segmenting no predefined clusters. This approach is used in situations where a training set of pre-classified records is unavailable (S. Y. Chen & Liu, 2004). Common tools for clustering are k-means and pattern-based clustering.
- c) Classification is the task of attempting to discover predictive patterns through classifying database records into a number of predefined categorical classes (S. Y. Chen & Liu, 2004).

8.3 Related Work

The use of a DSS in libraries has been in research for a long time. Scott Nicholson and Jeffrey Stanton (2006) present an overview of the projects documented in literature, some years before the concept of data mining and bibliomining became popularized. The authors conclude that these projects all shared a common focus on improving and automating two specific process of a library: acquisitions and collection management. An additional study by Maria Zamfir Bleyberg, Dongsheng Zhu, Karen Cole, Doug Bates and Wenyan Zhan (1999) describes the design and implementation of a prototype of a DSS for the Kansas State University libraries to provide information regarding patterns on the libraries' collection. The authors describe the proposed solution to incompatibilities emerged from the incorporation of external data to the model.

Several case studies describe different approaches to manage digital library collections based on DW techniques. For instance, the first concrete application of a DW to support management of information needs in a library is described by Joe Zucca (2003). S.C. Kao, H.C. Chang and C.H. Lin (2003) present a decision support tool for allocation of acquisition budget via data mining techniques. N. Girija and S.K. Srivatsa (2005) suggest the implementation of a DW to manage a digital library structure, they present a DW architecture model as well as a description of its major components. Chan-Chine Chang and Ruey-Shun Chen (2006) utilize bibliomining techniques, including data analysis, a DW and data mining to cluster readers' requirements. An additional framework for building a library DW is presented by Teh Ying Wah, Ng Hooi Peng and Ching Sue Hok (2007). The authors describe the main steps in the development of the library DW: data source, data extraction, data transformation and data loading, with special emphasis on the ETL functions and Web crawling for external data sources. Sun Lei and Chen Geng (2011) present the architecture of a DSS that combines the characteristics of a high-information system and of a DW. The purpose of this DSS is to analyze and mine data regarding collection management. Elaheh Homayounvala and Ammar Jalalimanesh (2012) propose a methodology to uncover patrons' research interests and promote research collaboration based on data mining techniques. The proposed methodology starts by creating a DW based on library operational data. Then, a knowledge map is created for library subjects and their usage. This knowledge map is analyzed to choose specific subjects for further analysis. Eventually, data mining techniques are applied to find subjects interpreted as research interests. In recent studies, Chen Bin (2013) describes the application of data mining tools in digital libraries management and Jishen Tang (2013) reports the implementation of a data mart to analyze the library readers' borrowing patterns. Mao Li Xu and Xiu Ying Li (2013) propose the architecture of a data warehouse-based DSS of university libraries for collection management. The authors focus on discussing the model design and implementation technology of the DW. A novel study by Xiu Mei Luan and He Jiang (2014) presents a system architecture of collection management based on a DW. The study uses OLAP tools to present the results, and provides an overview of the technologies utilized.

Only few studies are presented in literature regarding the use of bibliomining techniques as support tool for decision-making in services and collection. Reinhold Decker and Michael Hoppner (2006) describe a conceptual framework of a DSS for strategic planning and customer intelligence based on a DW structure. The authors recognize the importance of identifying relevant data structures and the challenges of leading with data heterogeneity. Markku Laitinen and Jarmo Saarti (2012) propose the use of a DW for improving the analysis of the library annual statistics of Finland. Finnish library statistics are based on indicators of the international standard ISO 11620 and some additional indicators developed for the particular needs of Finnish scientific libraries; these figures describe the library resources, library use and library collection. The authors

recognize that the current statistical database is not sufficient for decision-making, and that library management needs statistical data from other databases such as university output, publication impact factors and human resource's data. In most studies the authors concluded that the difficulties arise on deciding what data sources should be included, as well as on integrating the data coming from different platforms and applications.

8.4 Data Warehouse methodology selection

There are reasonably well-established approaches for implementing a DW; however, two classical methods are predominant: the one from Inmon and the one from Kimball. While both approaches obtain data from the same sources, they differ in the arrangement of these data in the DW itself (Lawyer & Chowdhury, 2004). The Inmon methodology, or the top-down approach, transfers the information from the various online transactions processing (OLTP) to a centralized DW, provided that this DW has some features: subject oriented, integrated, time variant and nonvolatile (Inmon, 2000, 2005). The Inmon approach, which follows a hub-and-spoke architecture, is characterized by the following: easy maintenance, normalized model, data integration of all areas, and complex design requiring high levels of expertise. Under this approach, the extracted and transformed data are typically stored in a third-normal form (3NF), from which data marts obtain their information. The Kimball methodology, on the other hand, is the union of smaller data marts, where every data mart represents a business process or dimensional mode (Kimball, 2006). This bottom-up approach is more iterative and modular, and prefers a high-level technical architecture also known as bus architecture (Ariyachandra & Watson, 2010). The Kimball approach is characterized by the following: business oriented process, dimensional model, data integration of business areas and short implementation time.

After analyzing both Inmon and Kimball methodologies, a hybrid approach, integrating the best of both methodologies, is adopted for this study. The DW architecture chosen for the iDSS implementation is Hefesto. The Hefesto approach, created by Ricardo D. Bernabeu in 2007, begins by collecting the information requirements of the user, followed by the extraction, transformation and loading of data in order to define the logical schema definition for the organization. Hefesto is characterized by the following: easy, realistic and simple to understand, is based on user requirements gathering, reduce the resistance to change, use conceptual and logical models, can be applied to DW and data marts, and is independent of technologies, physical structure and life cycle type (Bernabeu, 2010). The Hefesto methodology, as shown in Figure 8.3, considers four steps in the data warehouse creation process (Bernabeu, 2010). Firstly, a requirement analysis is performed; the user information needs are collected in order to define all queries of interest. Moreover, a set of resulting indicators are identified with their perspective of analysis, which are used to build the conceptual model of the DW. Secondly, the Hefesto methodology focuses on the analysis of OLTP systems, in order to determine how the indicators will be built, define correspondences and granularity, and build the extended conceptual model. Thirdly, the logical model is built, where the type of schema to be implemented is defined. Then the dimension and fact tables are drawn in order to create their respective unions. Finally, data are integrated by means of cleaning techniques, data quality control, and ETL processes. Policies and strategies for the initial loading of the DW are defined, as well as for its updating process.

8.5 iDSS for library holistic evaluation

A case study to demonstrate the applicability of bibliomining for library decision-making based on DW tools was performed in an Academic Library in Latin America (ALiLA). The analyzed library system includes a main library and two branches, housing collections numbering about 500,000 print volumes, multimedia content, and an array of digital resources. This academic library was selected for the case study because it belongs to one of the most prestigious universities in Latin America, and is considered one of the most modern and largest libraries of its country. A detailed

description of the analysis and design of the DW appears in Siguenza Guzman, L., Saquicela, V., and Cattrysse, D. (2014).

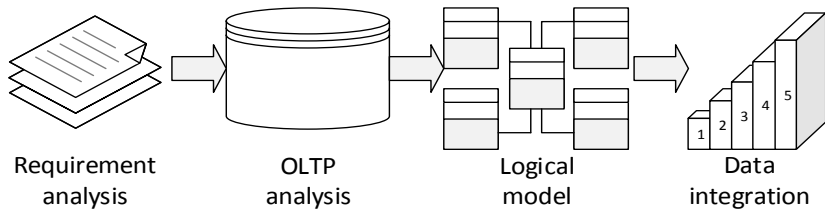


Figure 8.3: The Hefesto methodology

8.5.1 Architecture of the iDSS

Based on the theoretical DW architecture proposed by Siguenza-Guzman et al. (2015), and the methodology and technologies selected to develop the DW, the resulting DW architecture implemented for ALiLA is shown in Figure 8.4. The proposed iDSS architecture of ALiLA is structured in four blocks: 1) data collection, contains all sources used as data suppliers to the DW; 2) data extraction, cleansing and storage, selects and transforms the collected data; 3) data mining, is in charge of the application of data mining techniques; and 4) data presentation interprets and evaluates the results.

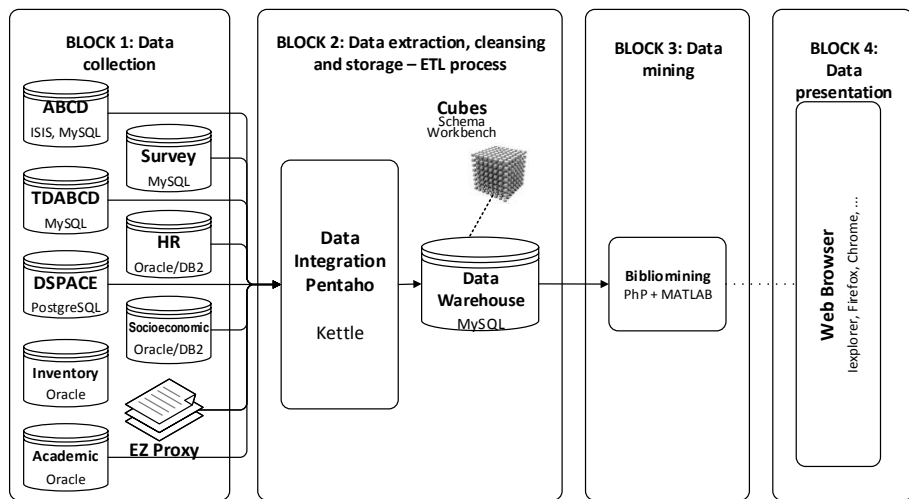


Figure 8.4: iDSS architecture of ALiLA

8.5.2 Design of the iDSS

The Hefesto methodology, as shown in Figure 8.3, consists of four steps. The first three steps deal with the *BLOCK 1: Data source* of the iDSS architecture (Figure 8.4), while the last step is focused on the *BLOCK 2: Data extraction, cleansing and storage*. Firstly, a requirement analysis to identify user information needs was performed. The requirements were analyzed through the four perspectives of the holistic evaluation framework proposed by Siguenza-Guzman et al. (2015). Based on the list of collected queries, the corresponding indicators and perspectives were defined.

Secondly, ten different ALiLA data sources were identified and analyzed based on the requirement analysis of the holistic evaluation approach. These data were generated at internal (4 data sources), university (4) and external level (2). Internal data sources refer to the databases that are managed at library level. University data sources are the databases managed at university level and exchange information flows with the library. External sources, in contrast, are not managed by internal processes of the library or the university; instead they are managed by companies like content management and database providers. Based on the requirement analysis and the list of data sources, a conceptual data model was also constructed.

Thirdly, a data mapping from the OLTP sources to the logical model was performed based on the conceptual data model. Moreover, the type of schema was defined. The data schema selected for the study was the star schema due to its simplicity and compatibility with the selected tools. In the star schema, facts are the core data elements being analyzed, and dimensions are the attributes about facts. By utilizing the star schema, the dimension and fact tables were built and joint to create multidimensional models.

Eventually, the data relevant for the holistic evaluation framework were integrated by using the technology tool Pentaho BI. Before loading the extracted data into the DW, a series of cleansing techniques and data quality controls had to be performed to transform and standardized the extracted data. For instance, due to a lack of uniform data entry, personal author's data were registered incorrectly; therefore string similarity measures like the Jaro-Winkler metric were used to indicate the percentage of similarity between fields. Once data were extracted, cleansed and transformed, they were loaded into de DW. A more detailed description of the analysis and design of the iDSS can be found in Siguenza Guzman, L., Saquicela, V., and Cattrysse, D. (2014).

8.6 Experimental use of bibliomining for library decision-making

The aim of data mining is to obtain knowledge from data to support decision-making. The knowledge discovery by means of data mining is related to the fact that a lot of information is unknown beforehand. Here, any suspicion can be tested about the behavior of different patterns within the collected data. Currently, an increasing interest in the use of data mining tools in libraries can be observed in literature; unfortunately none of these case studies covers the four library evaluation quadrants (Siguenza-Guzman, Saquicela, et al., 2015). In this section, the process applied for incorporating bibliomining techniques to the iDSS is presented. In particular, algorithms of regression, clustering and classification have been used in the case study for:

- a) *Predicting the future investment in library development.* This analysis utilizes the acquisition expenses in collection and technological infrastructure during the period 2000 – 2013 to forecast future investment. The parameters utilized are the acquisition value, and the corresponding month.
- b) *Clustering users for selective dissemination of information.* This analysis discovers group of users who share common interests and similar profiles, but belong to different faculties, in order to improve personalized recommendation
- c) *Predicting library factors that affect student cademic performance.* This analysis uses classification techniques to establish possible correlations between student library collection use and their final results taken on a pass/fail basis.

8.6.1 Predicting the future investment in library development

The predictive capabilities of data mining can be used to forecast library investments by means of regression techniques. For the case study, data from 2000 – 2013 are used to forecast one year of purchasing library collection and technological infrastructure. Three data mining techniques, namely, support vector machines (SVM), least-squares support vector machines (LSSVM) and

seasonal autoregressive integrated moving average (SARIMA) are compared in terms of root mean square error (RMSE) to select the appropriate technique for this dataset. SARIMA was utilized as a base line predictor, since it is a classical statistical technique to perform forecasting tasks (Box, Jenkins, & Reinsel, 2008). SVM and LSSVM were analyzed because data mining literature reports that these techniques provide good results for forecasting economic time series (Bouzerdoum, Mellit, & Massi Pavan, 2013; Yan & Chowdhury, 2013; Ahmad et al., 2014; Yan & Chowdhury, 2014; T.-T. Chen & Lee, 2015).

8.6.1.1 Data preparation

The attributes Acquisition Date and Value were analyzed using the acquisition multidimensional model defined in the DW for the first quadrant of the holistic matrix. The former indicates the month of acquisition, and the latter the value of the investment. To carry out the prediction analysis, 14 years of historical data were used, that is, the data of 168 months.

To create the time series representing the monthly expenses, data were grouped according to the month of acquisition. Firstly, outliers were detected by applying the median absolute deviation approach, as proposed by Iglewicz and Hoaglin (1993), and then, replaced by the mean values of the months where the outliers were detected. The rationale behind this is that a yearly seasonal effect can be observed in the time series as depicted in Figure 8.5a. These outliers represent atypical investments due to donations, political changes or law enforcements. For instance, thanks to a cooperation program with European universities, since 2008, ALiLA is investing on new technological infrastructure and information systems to improve its digital services. At the beginning of 2008, about \$70 thousand was invested on computers, multimedia equipment, bandwidth and a closed-circuit security system.

A noise reduction filter was then applied to the cleaned time series. The main objective of the filter is to smooth the signal in a way that the noise generated by unexpected events, such as last minute purchases and law enforcements are minimized. There is an extensive literature on the use of filters, being the hamming filter among the most commonly used and ease to implement (Priestley, 1983). The hamming filter is a low pass filter used to smoothen the signal by removing high frequency noise (Hamming, 1997). The resulting signal after applying the hamming filter is depicted in Figure 8.5b.

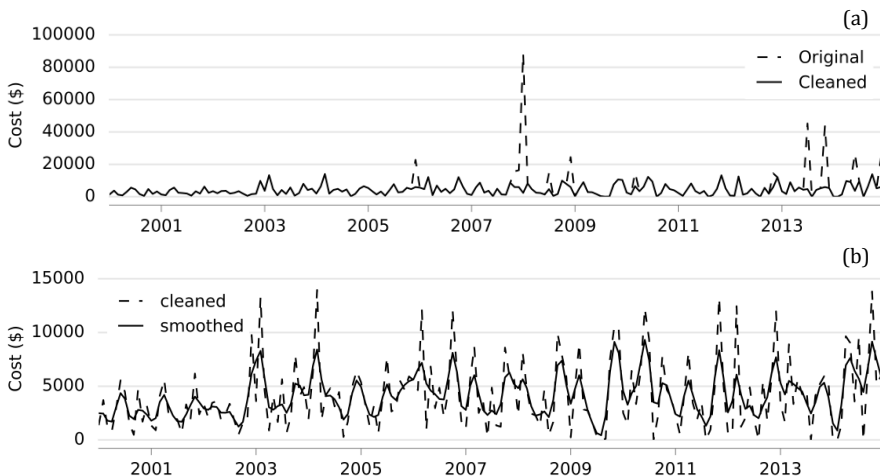


Figure 8.5: Acquisition values 2000-2014. (a) Original and cleaned time series after applying an outlier replacement process. (b) Cleaned and smoothed time series after applying a noise removal filter

8.6.1.2 Forecasting Library Investment

The objective in this task is to perform a 12-month forecast of investments in library collection and technological infrastructure. To do so, three data mining techniques, SVM, LSSVM and SARIMA, are first evaluated to select the more suitable for the dataset, by means of an out-of-sample test, as explained by Tashman (2000), which involves: (1) holdout split, (2) fitting the model, and (3) forecasting a year-ahead on the test period. The advantage of an out-of-sample test is that it mimics in a better way a real setting and provides a better estimation of the performance of forecasting models. The out-of-sample evaluation entails the division of the dataset in two parts, where the first part is used to train the model (fitting period) and the second part is used to evaluate the accuracy of the model (testing period) as depicted in Figure 8.6. The basic form of this process is known as fixed origin evaluation. The process of fixed origin evaluation focuses on a single origin, a good option in order to reduce computational effort; however, this univariate time series is very susceptible to be influenced by local characteristics. A possibility to overcome this problem is to update the origin recursively. A rolling-origin evaluation was selected to minimize this uncertainties related to predictions derived from a unique origin (Tashman, 2000). The rolling-origin evaluation starts by selecting a forecast origin T . This origin represents the end of the fitting period. Since the objective of the forecast task in this study is to predict the next 12 values of expenses in library collection and IT infrastructure, the forecast horizon H was set to 12. Then, rolling-origin evaluation is executed iteratively; and after every iteration, a new observation is added to the fitting period, i.e. the forecast origin T is shifted one month.

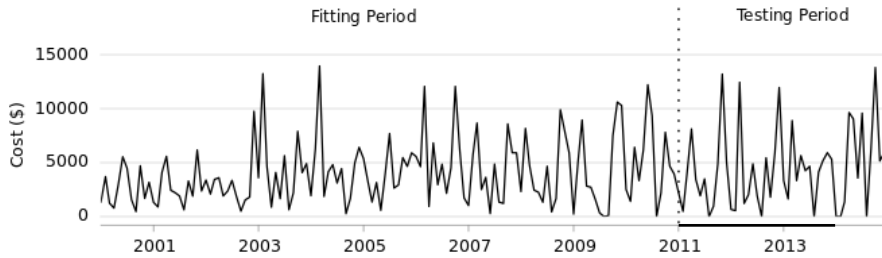


Figure 8.6: Out-of-sample evaluation

The dataset available for this task contains monthly data ranging from January 2000 to December 2014. The forecasting origin T was set to December 2009 on the first iteration. Thus, the fitting period, initially, comprises 120 months. The fitting period was used for the forecasting models to train and select the best hyper-parameters and coefficients. In the second iteration, the value of T is shifted one month, i.e. January 2010, increasing the fitting period by one month. Then again, the models are recalibrated. This decision was made to desensitize the errors produced by having a fixed T . The rolling-origin evaluation stops when the size of the testing period is equal to H , i.e. 12 months. At each iteration, the forecast models perform 12 predictions corresponding to a year-ahead forecast.

In the SARIMA model, the order was selected on the first iteration of the rolling-origin evaluation. The general notation for the order of a SARIMA model is given as $ARIMA(p, d, q) * (P, D, Q)_s$, in which the parameters inside the parentheses represent the order of (p) autoregression, (d) for differencing, and (q) for moving average in the model. The parentheses enclose the nonseasonal and the seasonal factors parameters, respectively. In this study, the resulting order of the SARIMA model was $(4,1,0)(2,1,0)$. This model was selected in a trial-and-error manner aiming to select a valid model with the smallest coefficients. In each iteration, the model coefficients are recalculated using the whole available fitting set.

The SVM and LSSVM are also trained and selected iteratively. As input features, these models utilized lagged values of the time series. This process is depicted in the equation below:

$$\begin{aligned}
Y_{t+1}^* &= f(y_t, y_{t-1}, \dots, y_{t-o+1}) \\
Y_{t+2}^* &= f(Y_{t+1}^*, y_t, y_{t-1}, \dots, y_{t-o+2}) \\
&\vdots \\
Y_{t+h}^* &= f(Y_{t+h-1}^*, \dots, y_t, \dots, y_{t-o+h})
\end{aligned}$$

where,

h	forecast horizon
o	order of the forecast model, i.e. the number of lagged values used as input features
$Y_{t+i}^*, 1 \leq i \leq h,$	predicted values
$y_{t-i}, 0 \leq i \leq o$	historical observations present in the fitting period.

Iteratively, the SVM and LSSVM models use the fitting period to select their hyperparameters and the optimal value of o . In SVM methods, parameter search plays a crucial role in the performance of the model. Some types of parameter search are employed in literature such v -fold cross-validation via parallel grid-search, heuristics search and inference of model parameters within the Bayesian evidence framework, being v -fold cross-validation the most reliable way for model parameter selection in median-sized problems (Olson & Delen, 2008). In v -fold cross-validation, the fitting period is first divided into v subsets. In the i th iteration ($i = 1, 2, \dots, v$), the i th set (validation set) is used to estimate the performance of the classifier trained on the remaining ($v-1$) sets (training set). The performance is generally evaluated by cost, e.g. the RMSE value. Considering the relatively small dataset currently available in this study, the hyperparameters are found by a grid-search process using 10-fold cross-validation for the following reasons: the cross-validation procedure can prevent an overfitting problem, and computational time to find good parameters by grid-search is not much more than other methods. This grid-search process is repeated using different configurations of o , i.e. $6 \leq o \leq 24, o \in \mathbb{Z}$. The model selected to perform the year-ahead forecasting on the testing period is the one that minimizes the RMSE value on the fitting period.

As result of the rolling approach, 36 year-ahead forecasts per model were obtained, as well as a descriptive distribution of the forecast errors. Figure 8.7 depicts the distribution of the resulting RMSE values and Table 8.1 shows the average RMSE values and their standard deviation per forecaster. A RMSE measurement has the same unit as the original data being simulated, and represents the sample standard deviation of the differences between predicted values and observed values. As can be seen, the three models behaved very similarly. LSSVM provides the lowest average RMSE. The SARIMA model has a similar RMSE value as LSSVM, however it produces the largest errors in the year-ahead forecast, and has the largest standard deviation. SVM produces the highest average RMSE, nevertheless its performance is similar to the other models. To determine whether the forecast results of the 12 month ahead prediction were statistically different, a Diebold-Mariano test (Diebold & Mariano, 1995) was performed. According to the two-tailed p -value, the forecast accuracies of the LSSVM and SVM model were not statistically significant ($p = 0.238$), neither the LSSVM and SARIMA forecast accuracies ($p = 0.631$).

For a better insight about the forecast predictions, Figure 8.8 shows the one step-ahead results of the testing period (2011-2014). The results show that the three methods were able to forecast the time series direction of change, e.g. increase or decrease trends; as well as, the magnitude of the predicted values. Although, the study shows promising results to forecast library investment, there is still room for improvements on the predictions, since time series depend on different exogenous factors such as sporadic donations, political changes and law enforcements that for this experimental exercise were not included in the predictive models. For instance, ALiLA library does not manage its own budget; instead, each faculty decides the amount to be invested in the library, based on their own finances, priorities and decisions. By incorporating variables that reflect these additional factors, improvements in the model can be expected.

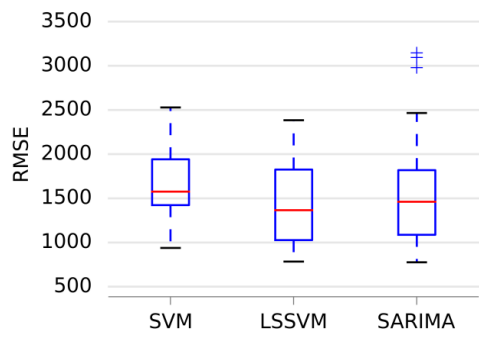


Figure 8.7: RMSE results of the three predictive models: (a) SVM, (b) LSSVM and (c) ARIMA

Table 8.1: Average RMSE results per forecaster

Forecaster	Average RMSE
LSSVM	1443.71 (+/- 880)
SARIMA	1577.74 (+/- 1250)
SVM	1638.68 (+/- 752)

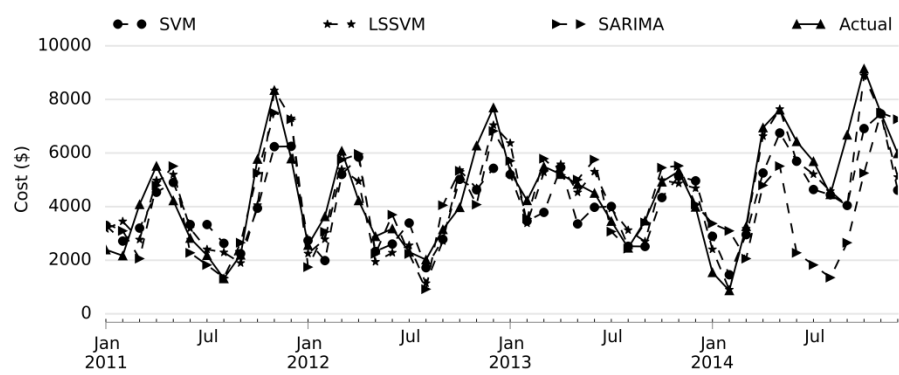


Figure 8.8: Comparison of the three techniques using step-ahead forecasting for the test period 2011 – 2014

To provide a better understanding of the forecasting results, Figure 8.9 depicts information regarding the behavior of the expenses per month and year. It can be observed that there is not a clear trend (almost constant) in the yearly expense values (Figure 8.9a), except for the last four years 2011-2014. but an intra-year pattern as depicted in Figure 8.9b. There are months with larger expenses such as March, October, November and December; furthermore, these months have the largest variability in the expense values with exception of November. Thus, the largest RMSE error values are found on these months. The rationale of these results is that large expenses occur at the beginning of each academic semester (March and October) and close to the end of the fiscal year (November, December).

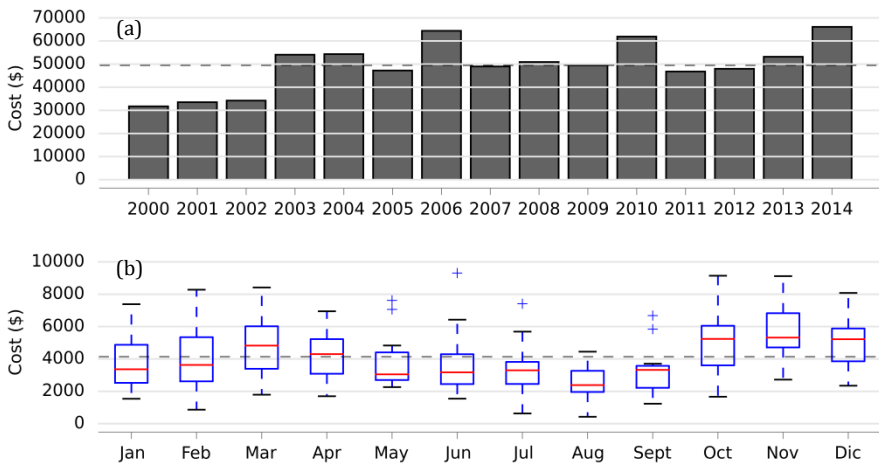


Figure 8.9: Expenses behavior: (a) per year; (b) per month. Horizontal dashed line represents the mean values

As above mentioned, the analysis and selection of the appropriate data mining technique for this dataset, as well as forecasting values can be improved by incorporating “exogenous variables”. An additional improvement to the model might be the introduction of additional data coming from other quadrants, such as analysis of information relevance (third quadrant) and usage indicators (fourth quadrant).

8.6.2 Users clustering for selective dissemination of information

Clustering is an unsupervised data mining technique to group similar elements according to a certain criterion; commonly, the referred criterion is a distance measurement between elements. In clustering problems, there is, often, no information regarding the correctness of the cluster relationships (Mitchell, Carbonell, & Michalski, 2011). In this work, a set of library users was clustered in order to find groups with similar profiles. User profiles are expressed in terms of demographic information and user behavior patterns in the library, i.e. categories and subcategories of the borrowed books, average loan duration, academic performance and socio-economic aspects, using knowledge area preferences as major considerations. An expected result of this exercise is to find at least as many clusters as faculties in the dataset, and that every cluster is composed by users of the same faculty. However, the real goal is to find clusters of users that share common interests and similar profiles, but belong to different faculties. These results can be used as input for the creation of library policies, as well as to increase recommendation accuracy to keep users with similar interests informed of new arrivals of library collection on specific topics.

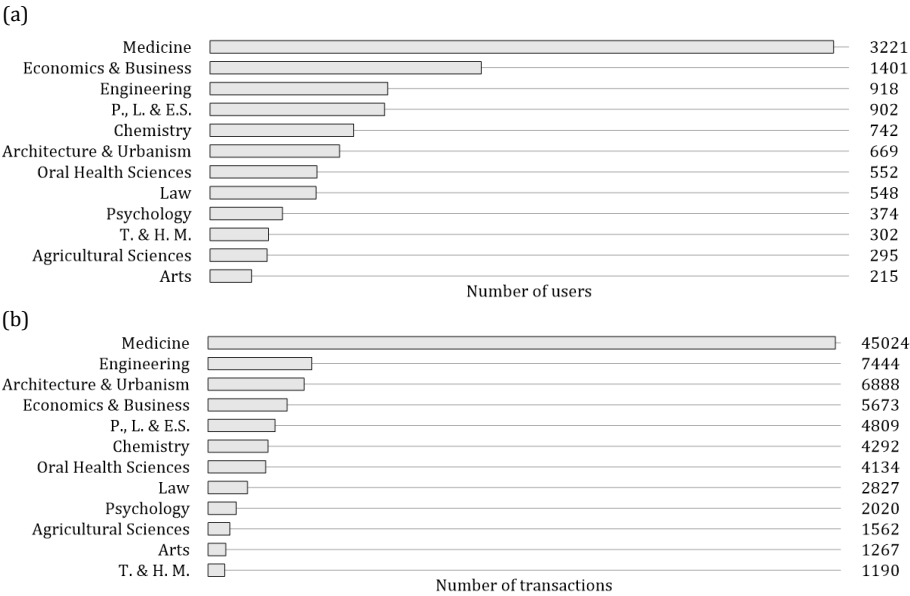
8.6.2.1 Data Preparation

The dataset used in this study was extracted from the data warehouse system implemented in ALiLA, three multidimensional models are consulted: academic, socioeconomic and lending. The available dataset contains information regarding the user profile and user behavior patterns, such as academic performance, economic income, loan durations and book category and subcategory (see Table 8.2). The relation between users and books is as follows, one user can loan many books, and a book can be borrowed by only one user at a time. The user profile regarding usage behavior was created by aggregating the book categories, and characterizing the loan duration with its mean and standard deviation.

Table 8.2: Variables used for clustering

Variable	Multidimensional model	Description
Faculty	Academic	Faculty to which the user belongs
School	Academic	Career to which the user belongs during that term
Academic performance	Academic	Academic performance rating: very good, good, average, and poor.
Household income	Socioeconomic	Average family income per month
Loan duration	Lending	Loan duration in hours
Category	Lending	Knowledge area of the book borrowed

Figure 8.10 depicts the number of users per faculty and the number of interactions with the university library. It is clear that the biggest faculties in terms of students and library transactions are Medicine and Economics & Business.



Note.- P. L.&E.S.=Philosophy, Language and Educational Sciences; T. & H. M.=Tourism and Hospitality Management

Figure 8.10: (a) Total number of users per faculty; (b) Number of user transactions per faculty

8.6.2.2 Clustering Library Users

The following clustering algorithms are tested: k-means, k-means++, k-means pca, mini-batch k-means, spectral clustering and ward. K-means is a prototype-based, popular old clustering algorithm that has been widely used due to its rapid processing ability (Wu, 2012). This algorithm attempts to find k non-overlapping clusters, in which k is chosen by the user. A variation of the traditional k-means is the k-means++ algorithm, designed to improve the centroid initialization for k-means (Vassilvitskii, 2007). The mini-batch k-means, a modified version of k-means, is considered faster than former-means and it is normally used for large datasets (Sculley, 2010). This algorithm attempts to optimize clustering results by taking mini-batches, to reduce computational time, as an input, which are random subsets of the whole dataset. K-means pca is also a variation of

the original k-means algorithm in which the cluster centroids are initialized according to the most explanatory variables in terms of variance; these values are found by applying a principal component analysis (PCA) on the original dataset; the assumption is that a strong relationship exists between the PCA subspace and the cluster centroids (Ding & He, 2004). Spectral clustering is a popular modern algorithm that is simple to implement and very often outperforms traditional clustering algorithms such as k-means (Luxburg, 2007). This algorithm relies on the eigenstructure of a matrix of point-to-point similarities (Bach & Jordan, 2004). Ward is an agglomerative hierarchical clustering algorithm that enables the clustering based on an objective similarity measure between clusters, referred to as the Ward distance (Mirkin, 2005).

Since all the clustering algorithms tested in this work require, as hyper-parameter, the number of resulting clusters, i.e. parameter n , 23 cluster configurations were proved for each algorithm, with parameter n ranging from 2 to 24. Since no information about the correctness of the clustering outputs was available, a validation measure, the silhouette score, was selected to evaluate the resulting clusters. This score is calculated for every cluster by comparing its tightness and separation; the final score is the combination of the silhouette values computed for every cluster (Rousseeuw, 1987). Well-clustered elements have a score near 1, while poorly-clustered elements have a score near -1. Figure 8.11 shows the silhouette scores for the different clustering algorithms. Spectral clustering and k-means pca underperform compared to the other algorithms. The remaining algorithms have similar silhouette scores; however the best scores were achieved by ward in cluster number 20, mini-batch k-means in cluster number 16 and spectral clustering in cluster number 2.

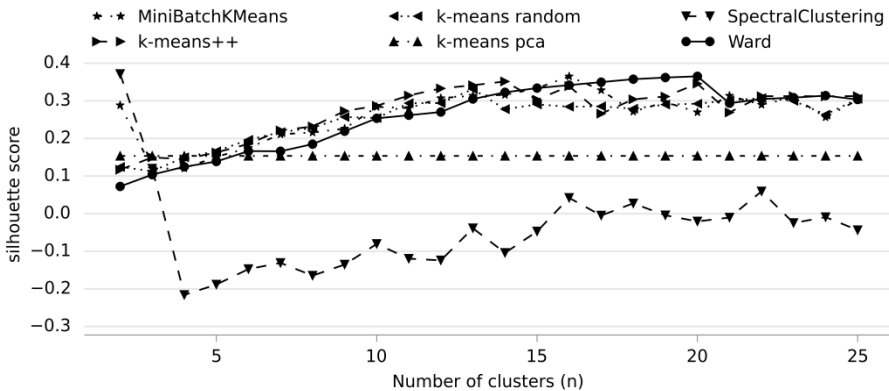
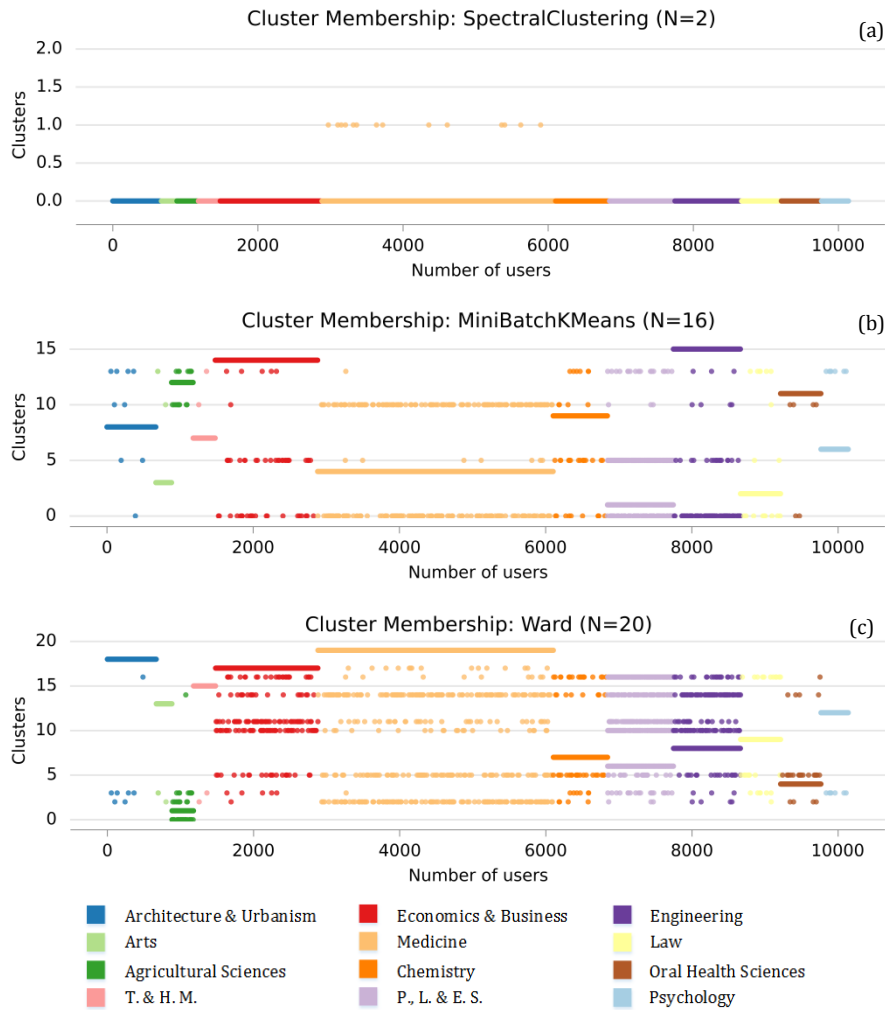


Figure 8.11: Silhouette score of the clustering algorithms at different configurations

Note.- P. L&E.S.=Philosophy, Language and Educational Sciences; T. & H. M.=Tourism and Hospitality Management

Figure 8.12 depicts the cluster membership of the best three clustering algorithms based on their silhouette score. In this figure, the x-axis represents library users identified by faculty membership (color). Each x-position represents a different library user. The y-axis represents the cluster allocation; each y value, starting from zero, represents a cluster. The intersection of library users and cluster allocation provides a visual interpretation of the cluster distribution for the different users.



Note.- P. L.&E.S.=Philosophy, Language and Educational Sciences; T. & H. M.=Tourism and Hospitality Management

Figure 8.12: Cluster membership grouped according to faculty of (a) spectral clustering (n=2); (b) mini-batch k-means (n=16); and, (c) ward (n=20)

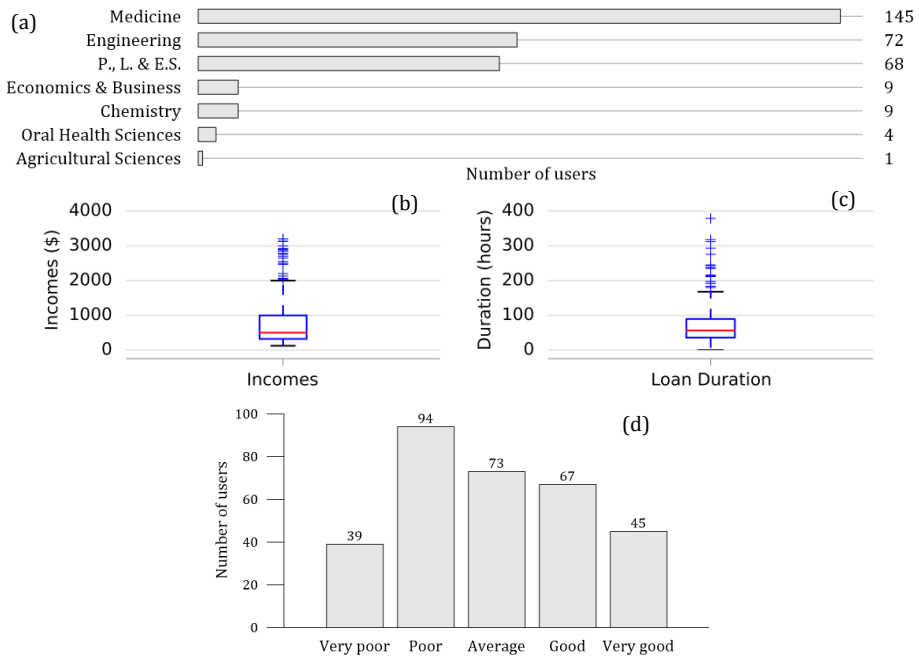
There are twelve faculties in the dataset; therefore it was expected to find, at least, one cluster per faculty. The clusters found by spectral clustering (n=2) are not useful, since the biggest cluster represents almost all the dataset. Mini-batch k-means (n=16) creates four clusters with users from different faculties; however these clusters contain just a few elements. For that reason, the clustering output found by ward (n=20) was selected as the most promising clustering output.

Since the goal of the clustering analysis is to find similar users from different faculties, we do not analyze clusters with users belonging to only one faculty. Clusters with users of the same faculty are natural user divisions; therefore they do not provide opportunities to find interesting new patterns. An intra-cluster analysis was performed to better understand the existing pattern in a given cluster.

To demonstrate the use of clustering techniques in this dataset, the intra-cluster analysis focuses on the clustering results obtained by ward ($n=20$), specifically cluster number 14. As can be seen in Note.- P. L.&E.S.=Philosophy, Language and Educational Sciences; T. & H. M.=Tourism and Hospitality Management

Figure 8.12, the referred cluster consists of users from different faculties, and where the share of users is representative.

For the intra-cluster analysis, firstly cluster statistics are analyzed. As can be seen in Figure 8.13a, users mostly come from the faculties of Medicine, Engineering, and Philosophy, Language and Educational Sciences. This, in fact, is an interesting cluster since users have very different backgrounds. Figure 8.13b-c provide information about the distribution in terms of income and loan duration. These two figures do not surprise, as they report average values, i.e. average income and average loan duration, even when compared to other clusters. Finally, Figure 8.13d shows the academic performance of users that belong to this cluster. As can be seen, the academic performance of the most of users is rather low.



Note.- P. L.&E.S.=Philosophy, Language and Educational Sciences

Figure 8.13: Statistics of the cluster 14 produced by ward ($n=20$): a) number of users per faculty; b) income; c) loan duration; d) academic performance

The final part of the performed clustering analysis is to find patterns in the topics of interest based on the books borrowed by users. Every book in the library has associated one category and subcategory. In ALiLA, this classification is given by the Dewey Decimal Classification (DDC). DDC is the most widely used classification system, it divides knowledge disciplines or fields of study into ten main classes, as follows: 000 Computer science, information and general works, 100 Philosophy and psychology, 200 Religion, 300 Social sciences, 400 Language, 500 Science, 600 Technology, 700 Arts & recreation, 800 Literature, and 900 History and geography (Dewey, 2011).

Figure 8.14a-b depicts the categories and subcategories that the clustered users were interested in. For a better understanding, the structure of the plots shown in Figure 8.14 can be seen as a matrix whose cells can have only binary values, i.e. present (color), absent (white). Every column represents a user's interest in certain categories or subcategories. Finally, every row represents one category or subcategory. If a cell has a white color background, the corresponding user was not interested in the corresponding category or subcategory. In the analyzed cluster, all users were interested in the Computer science category. Additionally, users were also interested in Social sciences and Science. To be more precise, Figure 8.14b provides information about the subcategories where users have common interests. Here, the dominant subcategory is Computer science, information & general works, which, according to the Dewey system, entails Computer science topics and programming languages. The second dominant category is Medicine. Even though most of the users in this cluster come from the faculty of Medicine, this result is still interesting, since the distribution of that subcategory is across users from different faculties. To a lesser extent, subcategories such as Mathematics, Biology and Social sciences are also present across the users in this cluster; however, their distributions are sparser.



† CS, I & GW = Computer science, information and general works

Figure 8.14: User behavior in terms of topics of interest per faculty for cluster 14, ward (n=20); (a) user interests in terms of categories; (b) user interests in terms of subcategories

The analysis of cluster 14 produced by ward (n=20) shows that their users are heavily interested by technical subjects. This type of clusters can be used as input for a more in-depth analysis in order to ensemble, for instance, a multidisciplinary team to perform research activities.

8.6.3 Predicting factors that affect student academic performance

Once again, the predictive capabilities of data mining can be used now to classify student academic performance based in several factors such as library usage, academic and socio-economic profiles by means of a classification task. In machine learning, classification is a supervised task aimed to be used as a mapping function (classifier) of multi-dimensional observations into a finite number of output classes. The classifier is created from labeled data, i.e. training data, to maximize the correct assignation of predicted labels and ground truth. For the sake of simplicity, the academic performance is evaluated based on a pass/fail basis, and thus characterized as a binary problem. Even though there are different academic performance scales, such as very poor, poor, average, good and very good, a binary problem seems more suitable for the available dataset, since the academic performance depends on many factors like academic teaching skills, students' attitudes and aptitude, and student attendance (Cox & Jantti, 2012), which are not considered in this study.

8.6.3.1 Data preparation

The same as for clustering, the dataset used in this exercise was extracted from the data warehouse system implemented in ALiLA, three multidimensional models are consulted: academic, socioeconomic and lending. The available dataset contains information regarding the user profile and user behavior patterns, such as academic performance, economic income, loan durations and book category and subcategory (Table 8.3). For the classification process, data from the academic year September 2013 – August 2014 are used, that is semester September 2013 – February 2014 and semester March 2014 – August 2014. In addition, users, who have less than three library transactions, are filtered out.

Table 8.3: Variables used for classification

Variable	Multidimensional model	Description
Faculty	Academic	Faculty to which the user belongs
School	Academic	Career to which the user belongs during that term
Household income	Socioeconomic	Average family income per month
Loan duration	Lending	Loan duration in hours
Category	Lending	Knowledge area of the book borrowed
Subcategory	Lending	Subcategory knowledge area of the book borrowed

As mentioned before, the output variable for the classification model is the academic performance. This is a binary variable, which was created by merging the states very poor and poor as class fail, and average, good and very good as class pass. The input variables were standardized in a way that every variable has zero mean and unit variance.

As in the forecasting exercise, a SVM model was selected to test the feasibility and accuracy of the model given the input data. The first analysis was performed on the whole dataset, which includes all faculties and careers, and the second analysis was performed on the "Medicine and Surgery" career, since it has the most number of users.

8.6.3.2 Impact of libraries in academic performance

The analysis of library students, in terms of loan duration and income, shows that there is not a clear difference in their statistical distributions (Figure 8.15a, and Figure 8.15b). This provides evidence that these variables are not likely to be well suited to separate the output classes on the model. In addition, Figure 8.15c shows that, in general, there are 34.32% more students passing than students failing. Therefore, this is an unbalanced binary classification problem.

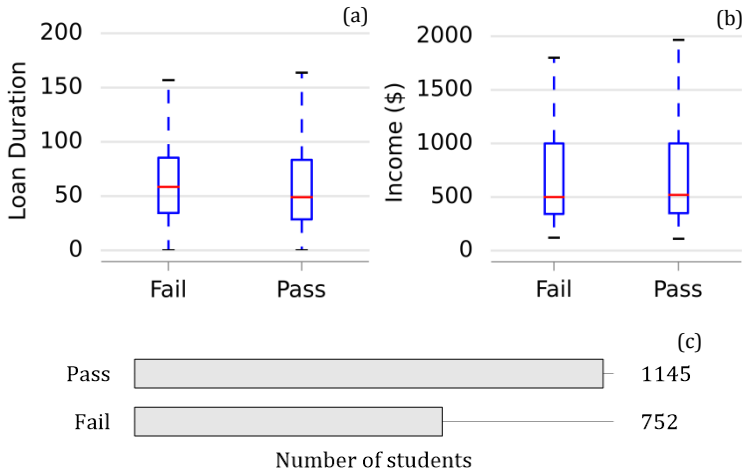


Figure 8.15: Statistics of the full dataset: (a) income; (b) loan duration; (c) academic performance

A correlation analysis, by means of the Pearson coefficient, was performed on the input variables with respect to the output variable. As can be seen in Table 8.4, just few variables had a correlation value r different than zero, and at the same time, statistically significant. The largest correlation was found on the variable career ($r = 0.104$). This supports the idea that every career has its own methodology to grade students.

Table 8.4: Correlation of variables including their p-value

Variable	r	p-value
n_transactions	0.050	0.029
income	0.081	0.000
avg_loan	-0.075	0.001
std_loan	-0.064	0.005
technology	0.046	0.047
career	0.104	0.000

In order to exhibit, in a graphical form, how separable the two output classes are in the classification problem, two plots, i.e. parallel coordinates (Heinrich & Weiskopf, 2012) and Andrews curves (Andrews, 1972), are shown in Figure 8.16. The parallel coordinates plot aims to visualize high dimensional data. For this work, the coordinates chosen were the ones with the highest correlation factors. Another useful plot to visualize high dimensional data is Andrews' curves plot. This plot is aimed to reveal structure on the dataset; for this purpose, every data point $x_i = (x_1, \dots, x_k)$ with dimension k defines a finite Fourier series:

$$f_x(t) = \frac{x_1}{\sqrt{2}} + x_2 \sin(t) + x_3 \cos(t) + x_4 \sin(2t) + x_5 \cos(2t) + \dots$$

This function is plotted on the range $-\pi < t < \pi$. Thus, every data point is represented as a line in the mentioned range. When the observations represented by the input variables are easy to separate into output classes, their colored lines follow a pattern such that, visually, their separation becomes trivial. As can be seen in Figure 8.16, there is a very limited structure on the dataset. This reinforces the idea that the separability between the two classes is not trivial and, therefore, non-linear.

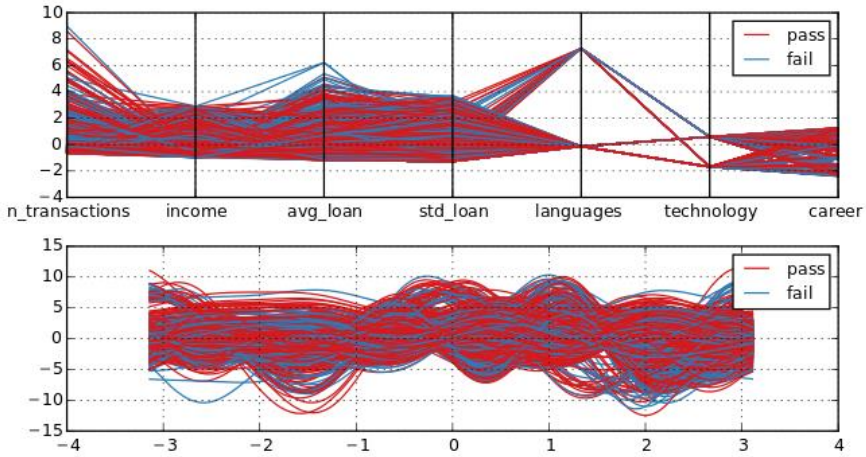


Figure 8.16: High dimensional visualization of the complete dataset. (a) Parallel coordinates. (b) Andrew's curves

In the training phase, the hyper-parameters were found by a ten-fold cross-validation on the training set. Then, the quality assessment was performed on the testing dataset. In this phase, 70% of the dataset was used for training, and the rest for testing. The hyper-parameters were found by a grid search process. Additionally, since the two classes, i.e. pass and fail, are unbalanced, the class weights were computed according to the number of elements on each class.

The first approach to train the model was to find the SVM hyper-parameters by maximizing the accuracy $\left(\text{Acc} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}\right)$, where T_p is the amount of instances of class fail correctly classified, T_n is similar to T_p but for the class pass. F_p represents the amount of instances classified as class fail that actually belongs to class pass. F_n is similar to F_p but for class pass. This is the classical approach in classification problems; however it was not possible to obtain accuracy higher than 59% on the test set. For this reason, we then focused on the single classes in order to have good predictive results. After some experiments, we found that it is possible to have good results on the class pass by training the algorithm to maximize the precision on this class. The precision is given by $\left(\text{Precision} = \frac{T_p}{T_p + F_p}\right)$.

The accuracy on the training set was equal to 66% (+/-0.09). For a better insight on the prediction results, Table 8.5 shows the most important classification metrics on the training set. Here, the precision on the class pass is the most relevant result (71%). The accuracy on the testing set was equal to 59%, and the remaining classification metrics are shown in Table 8.6. On the test set, the precision on the class pass was 69%. Since the classification metrics are similar between the training and testing datasets, it is possible to state that the training phase was performed correctly and there were not overfitting issues. The classification results on the testing dataset suggests that the model is not suitable to predict accurately whether a library user will pass or not given its user profile and information regarding library usage. Another important aspect to stress is that the model is able to find only 61% (recall) of the users that pass, however there is a high certainty (69% precision) that the library users actually passes. In other words, the model is able to find a representative fraction of library users that pass with a relative high certainty. For a complete view of the classification results, Table 8.7 shows the confusion matrix of the testing set.

Table 8.5: Classification results for the training set

Variable	Precision	Recall	F1-score	Support
Fail	0.54	0.62	0.58	537
Pass	0.71	0.64	0.67	790
Avg/Total	0.64	0.63	0.63	1327

Table 8.6: Classification results for the testing set

Variable	Precision	Recall	F1-score	Support
Fail	0.46	0.55	0.50	215
Pass	0.69	0.61	0.65	355
Avg/Total	0.60	0.59	0.59	570

Table 8.7: Confusion Matrix for the binary classification model on the testing set

Actual/Predicted	Fail	Pass
Fail	118	97
Pass	139	216

The final classification exercise was performed on the Medicine career, since this is the career with the most significant number of library users. As shown in Figure 8.17a, the distribution of loan durations of both classes is very similar. This does not hold when comparing the distributions of incomes (Figure 8.17b). The library users with higher income have more chances to pass. Furthermore, the classes are imbalanced; there are 50.7% more library users that passes than fails, as depicted in Figure 8.17c.

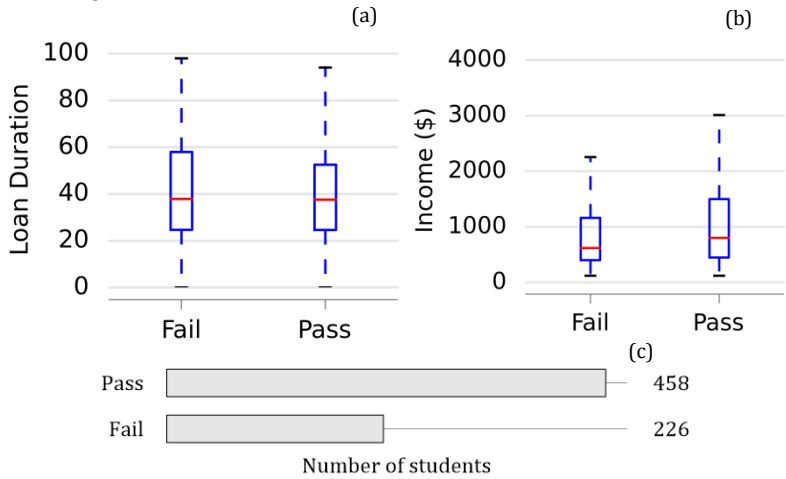


Figure 8.17: Statistics of the career of Medicine and Surgery: a) income; b) loan duration; c) academic performance

The correlation analysis on the input variables (Table 8.8) show similar results as the ones found by using the whole dataset (Table 8.4); however, there is a very low correlation and, in some cases, small statistical significance.

Table 8.8: Correlation of variables including their p-value

Variable	r	p-value
income	0.093	0.015
languages	0.013	0.732
std_loan	0.011	0.779
n_transactions	0.009	0.822
avg_loan	-0.028	0.460
technology	-0.067	0.079

A graphical representation of the dataset is shown in Figure 8.18. As in the case of the full dataset, it is not possible to find a clear structure on the dataset.

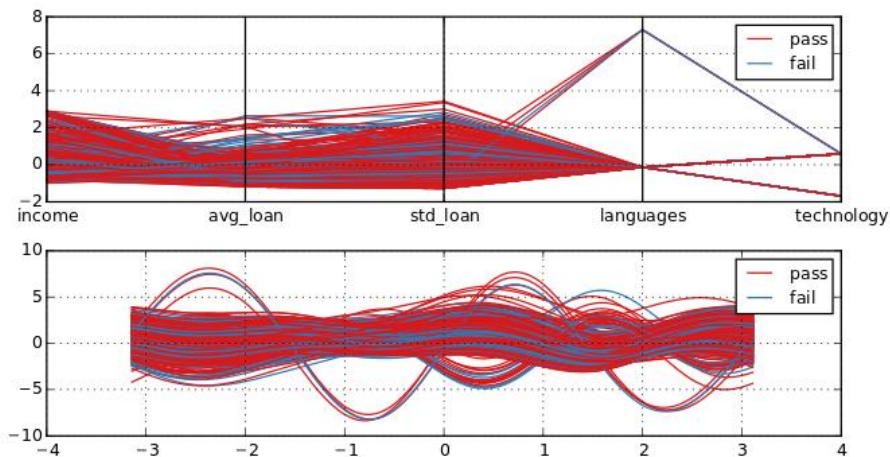


Figure 8.18: High dimensional visualization of the Medicine dataset. (a) parallel coordinates. (b) Andrews curves

The classification phase was performed with the same methodology as explained for the complete dataset. The model was trained to maximize the precision on the class pass; since, as in the previous case, it is more likely to have better classification results. Similarly, the training period was performed taking into account the unbalance of classes; consequently, giving different class weights to them.

The accuracy on the training set was 58% (+/- 0.14), and for the testing set 50%. Table 8.9 shows the classification metrics for the two classes on the test set. The model was able to find 47% of the library users that actually passed (recall) with a certainty of 76% (precision). These results outperform the ones found by using the whole dataset, since the precision is higher, i.e. there is more certainty that library users predicted to pass, in fact passed. Finally, the confusion matrix of the classification results on the test set is shown in Table 8.10.

Table 8.9: Classification results for the testing set

Variable	Precision	Recall	F1-score	Support
Fail	0.29	0.58	0.38	55
Pass	0.76	0.47	0.58	151
Avg/Total	0.63	0.50	0.53	206

Table 8.10: Confusion Matrix for the Binary Classification Model on the training set

Actual/Predicted	Fail	Pass
Fail	32	23
Pass	80	71

No notable associations were found among socioeconomic background, library usage, and academic performance; although, family income factors may affect student academic performance, since it was found that library users with higher income have more chances to pass. For the institution, these types of findings can support the development of library services to target specific student groups on the basis that higher library usage may lead to improved academic performance. In addition, this experimental study describes a research design that is replicable in other libraries and contributes to library usage and learning analytics literature.

The preliminary findings of this experimental study provide basis for further investigation on this topic and demonstrate how institutional data can be combined to examine library usage and academic performance at a single institution. Undergraduate students and library usage data were analyzed to identify results that suggested associations or relations between library usage and academic performance.

8.7 Conclusions

Based on a holistic evaluation framework for data collection and an integrated decision support system based on a DW architecture implemented in an academic library in Latin America, this study presents an experimental use of data mining techniques for library decision-making. To this end, algorithms of regression, clustering and classification have been applied to the case study in the following manner: 1) predicting the future investment in library development; 2) finding clusters of users that share common interests and similar profiles, but belong to different faculties; 3) predicting library factors that affect student academic performance by analyzing possible correlations of library usage and academic performance. Despite its experimental character, the reported research shows that the use of data mining techniques utilizing as a basis a holistic approach for data input, can provide the data-based justification for identified library needs, or the appropriateness of particular decisions. The most important implication of this exercise is that having as a baseline the holistic framework and the iDSS architecture, the study demonstrates the potential to apply data mining techniques to the integrated framework in order to understand the entire library system. There are, however, some considerations to be taken into account, mainly regarding the quality and effectiveness of the results, which depend to a great extent on the availability of relevant information and the quality of data... the more data are provided; the more accurate results will be obtained.

In the future, recommendations improving the accuracy of the results, implementing more advanced data mining techniques such as association algorithms, or tackling other quadrants of the holistic evaluation framework should be considered.

References

- Ahmad, A. S., Hassan, M. Y., Abdullah, M. P., Rahman, H. A., Hussin, F., Abdullah, H., & Saidur, R. (2014). A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renewable and Sustainable Energy Reviews*, 33, 102–109. <http://doi.org/10.1016/j.rser.2014.01.069>
- Andrews, D. F. (1972). Plots of High-Dimensional Data. *Biometrics*, 28(1), 125–136. <http://doi.org/10.2307/2528964>
- Ariyachandra, T., & Watson, H. (2010). Key organizational factors in data warehouse architecture selection. *Decision Support Systems*, 49(2), 200–212. <http://doi.org/10.1016/j.dss.2010.02.006>

- Bach, F. R., & Jordan, M. I. (2004). Learning spectral clustering. In S. Thrun & L. K. Saul (Eds.), *Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference* (pp. 305–312). MIT Press.
- Bernabeu, R. D. (2010). *Data Warehousing: Research and Concept Systematization - HEFESTO: Methodology for the Construction of a Data Warehouse*. Cordova, Argentina. Retrieved from <http://www.businessintelligence.info/docs/hefesto-v2.pdf>
- Bin, C. (2013). Study on Data Mining in Digital Libraries. In Y. Yang, M. Ma, & B. Liu (Eds.), *Information Computing and Applications* (pp. 282–291). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-53703-5_30
- Bleyberg, M. Z., Zhu, D., Cole, K., Bates, D., & Zhan, W. (1999). Developing an integrated library decision support data warehouse. In 1999 IEEE International Conference on Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings (Vol. 2, pp. 546–551). Tokyo, Japan: IEEE. <http://doi.org/10.1109/ICSMC.1999.825319>
- Bouzerdoun, M., Mellit, A., & Massi Pavan, A. (2013). A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. *Solar Energy*, 98, Part C, 226–235. <http://doi.org/10.1016/j.solener.2013.10.002>
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control* (4 edition). Hoboken, N.J: Wiley.
- Chang, C.-C., & Chen, R.-S. (2006). Using data mining technology to solve classification problems: A case study of campus digital library. *Electronic Library*, The, 24(3), 307–321. <http://doi.org/10.1108/02640470610671178>
- Chen, S. Y., & Liu, X. (2004). The contribution of data mining to information science. *Journal of Information Science*, 30(6), 550–558. <http://doi.org/10.1177/0165551504047928>
- Chen, T.-T., & Lee, S.-J. (2015). A weighted LS-SVM based learning system for time series forecasting. *Information Sciences*, 299, 99–116. <http://doi.org/10.1016/j.ins.2014.12.031>
- Cox, B., & Jantti, M. (2012). Discovering the impact of library use and student performance. *Deputy Vice-Chancellor (Academic) - Papers*, 1–9.
- Curtis, M. B., & Joshi, K. (2011). Developing a Data Warehouse: Some Guidelines and Suggestions. *The Review of Business Information Systems*, 3(2), 13–26.
- Decker, R., & Höppner, M. (2006). Strategic planning and customer intelligence in academic libraries. *Library Hi Tech*, 24(4), 504–514. <http://doi.org/10.1108/07378830610715374>
- Dewey, M. (2011). Dewey Decimal Classification, DDC 23. OCLC Online Computer Library Center.
- Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13(3), 253–265.
- Ding, C., & He, X. (2004). K-means Clustering via Principal Component Analysis. In *Proceedings of the Twenty-first International Conference on Machine Learning* (p. 29–). New York, NY, USA: ACM. <http://doi.org/10.1145/1015330.1015408>
- Girija, N., & Srivatsa, S. K. (2005). Constructing the virtual library data warehouse from a blueprint. *Information Technology Journal*, 4(3), 246–250.
- Hamming, R. W. (1997). *Digital Filters* (3 edition). Mineola, N.Y: Dover Publications.
- Heinrich, J., & Weiskopf, D. (2012). State of the Art of Parallel Coordinates. In M. Sbert & L. Szirmay-Kalos (Eds.), *Eurographics 2013 - State of the Art Reports*. The Eurographics Association. <http://doi.org/10.2312/conf/EG2013/stars/095-116>
- Hobohm, H.-C. (2004). *Knowledge Management: Libraries and Librarians Taking Up the Challenge*. Walter de Gruyter.
- Homayounvala, E., & Jalalimanesh, A. (2012). Promoting Research Collaboration Based on Data Mining Techniques in Library Information Systems. *International Journal of Information Technology and Business Management*, 8(1), 73–82.
- Iglewicz, B., & Hoaglin, D. C. (1993). *How to detect and handle outliers* (Vol. 16). Milwaukee, Wis.: ASQC Quality Press.
- Inmon, W. H. (2000). *What is a Data Warehouse?*
- Inmon, W. H. (2005). *Building the data warehouse* (4 ed.). John Wiley & Sons.
- Kao, S. C., Chang, H. C., & Lin, C. H. (2003). Decision support for the academic library acquisition budget allocation via circulation database mining. *Information Processing & Management*, 39(1), 133–147. [http://doi.org/10.1016/S0306-4573\(02\)00019-5](http://doi.org/10.1016/S0306-4573(02)00019-5)
- Kimball, R. (2006). *The data warehouse toolkit*. John Wiley & Sons.

- Lai, M.-C., Wang, W.-K., Huang, H.-C., & Kao, M.-C. (2011). Linking the benchmarking tool to a knowledge-based system for performance improvement. *Expert Systems with Applications*, 38(8), 10579–10586. <http://doi.org/10.1016/j.eswa.2011.02.101>
- Laitinen, M., & Saarti, J. (2012). A model for a library-management toolbox: Data warehousing as a tool for filtering and analyzing statistical information from multiple sources. *Library Management*, 33(4/5), 253–260. <http://doi.org/10.1108/01435121211242290>
- Lawyer, J., & Chowdhury, S. (2004). Best practices in data warehousing to support business initiatives and needs. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, 2004 (p. 9 pp.-). <http://doi.org/10.1109/HICSS.2004.1265515>
- Lei, S., & Geng, C. (2011). Design and Implementation of Library Information Decision Support System. In *Management and Service Science (MASS)*, 2011 International Conference on (pp. 1–4). <http://doi.org/10.1109/ICMSS.2011.5998163>
- Luan, X. M., & Jiang, H. (2014). Design of Books Analysis System in University Library Based on Data Warehouse. *Applied Mechanics and Materials*, 513-517, 2121–2124. <http://doi.org/10.4028/www.scientific.net/AMM.513-517.2121>
- Luxburg, U. von. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17(4), 395–416. <http://doi.org/10.1007/s11222-007-9033-z>
- Mirkin, B. (2005). *Clustering for Data Mining: A Data Recovery Approach*. CRC Press.
- Mitchell, T. M., Carbonell, J. G., & Michalski, R. S. (Eds.). (2011). *Machine Learning: A Guide to Current Research* (Softcover reprint of the original 1st ed. 1986 edition). Springer.
- Nicholson, S. (2003). The bibliomining process: Data Warehousing and Data Mining for library decision-making. *Information Technology and Libraries*, 22(4), 146–151.
- Nicholson, S. (2004). A conceptual framework for the holistic measurement and cumulative evaluation of library services. *Journal of Documentation*, 60(2), 164–182. <http://doi.org/10.1108/00220410410522043>
- Nicholson, S., & Stanton, J. (2006). Bibliomining for library decision-making. In *Encyclopedia of Data Warehousing and Mining* (Second Edition, pp. 100–105). Retrieved from <http://www.igi-global.com/chapter/encyclopedia-data-warehousing-mining/10591>
- Nicholson, S., & Stanton, J. (2009). Bibliomining for library decision-making. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (Second Edition, pp. 341–345). IGI Global. Retrieved from <http://www.igi-global.com/chapter/encyclopedia-information-science-technology-second/13596>
- Nicholson, S., & Stanton, J. M. (2003). Gaining strategic advantage through Bibliomining: Data Mining for management decisions in corporate, special, digital, and traditional libraries. In H. R. Nemati & C. D. Barko, *Organizational data mining: Leveraging enterprise data resources for optimal performance* (pp. 247–262). Hershey, PA, USA: Idea Group Publishing (an imprint of Idea Group Inc.). Retrieved from http://arizona.openrepository.com/arizona/bitstream/10150/106383/1/Nicholson_3.pdf
- Olson, D. L., & Delen, D. (2008). *Advanced Data Mining Techniques*. Springer Science & Business Media.
- Radescu, L. C. (2003). Managing a Data Warehouse Environment. *Annals. Computer Science Series*, 1(2), 82–93.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [http://doi.org/10.1016/0377-0427\(87\)90125-7](http://doi.org/10.1016/0377-0427(87)90125-7)
- Sculley, D. (2010). Web-scale K-means Clustering. In *Proceedings of the 19th International Conference on World Wide Web* (pp. 1177–1178). New York, NY, USA: ACM. <http://doi.org/10.1145/1772690.1772862>
- Siguenza-Guzman, L., Holans, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2013). Towards a holistic analysis tool to support decision-making in libraries. In *Proceedings of the IATUL Conferences* (p. Paper 29). Cape Town, South Africa: Purdue e-Pubs. Retrieved from <http://docs.lib.purdue.edu/iatul/2013/papers/29/>
- Siguenza-Guzman, L., Saquicela, V., Avila, E., Vandewalle, J., & Cattrysse, D. (2015). Literature review of data mining applications in academic libraries. Accepted to be published at the *Journal of Academic Librarianship*.

- Siguenza-Guzman, L., Saquicela, V., & Cattrysse, D. (2014). Design of an Integrated Decision Support System for library holistic evaluation. In *Proceedings of IATUL Conferences* (pp. 1–12). Espoo, Finland.
- Siguenza-Guzman, L., Van Den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2015). A holistic approach to supporting academic libraries in resource allocation processes. *The Library Quarterly: Information, Community, Policy*, 85(3).
- Tang, J. (2013). Study of Analysis Data Mart in Library Borrowing. In Y. Yang & M. Ma (Eds.), *Proceedings of the 2nd International Conference on Green Communications and Networks 2012 (GCN 2012): Volume 4* (Vol. 4, pp. 671–678). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-35440-3_86
- Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4), 437–450. [http://doi.org/10.1016/S0169-2070\(00\)00065-0](http://doi.org/10.1016/S0169-2070(00)00065-0)
- Townley, C. T. (2001). Knowledge Management and Academic Libraries. *College & Research Libraries*, 62(1), 44–55.
- Vassilvitskii, S. (2007). K-means: Algorithms, Analyses, Experiments. Stanford University.
- Wrembel, R., & Koncilia, C. (2007). *Data Warehouses and OLAP: Concepts, Architectures, and Solutions*. Idea Group Inc. (IGI).
- Wu, J. (2012). *Advances in K-means Clustering: A Data Mining Thinking*. Springer Science & Business Media.
- Xu, M. L., & Li, X. Y. (2013). Construction of the Library Management System Based on Data Warehouse and OLAP. *Applied Mechanics and Materials*, 380-384, 4796–4799. <http://doi.org/10.4028/www.scientific.net/AMM.380-384.4796>
- Yan, X., & Chowdhury, N. A. (2013). Mid-term electricity market clearing price forecasting: A hybrid LSSVM and ARMAX approach. *International Journal of Electrical Power & Energy Systems*, 53, 20–26. <http://doi.org/10.1016/j.ijepes.2013.04.006>
- Yan, X., & Chowdhury, N. A. (2014). Mid-term electricity market clearing price forecasting: A multiple SVM approach. *International Journal of Electrical Power & Energy Systems*, 58, 206–214. <http://doi.org/10.1016/j.ijepes.2014.01.023>
- Ying Wah, T., Hooi Peng, N., & Sue Hok, C. (2007). Building Data Warehouse. In *Proceedings of the 24th South East Asia Regional Computer Conference*. Bangkok, Thailand. Retrieved from http://www.ijcim.au.edu/SpecialEditions/v15nSP4/P05SEARCC_BuildingDataWarehou e.pdf
- Zhang, Y. (2010). Developing a holistic model for digital library evaluation. *Journal of the American Society for Information Science and Technology*, 61(1), 88–110. <http://doi.org/10.1002/asi.21220>
- Zucca, J. (2003). Traces in the clickstream: Early work on a management information repository at the University of Pennsylvania. *Information Technology and Libraries*, 22(4), 175–178.

Chapter 9: **Towards an optimization**

model for resource allocation in

libraries

Siguenza-Guzman, L., Vanegas, P., V., & Cattryse, D. Towards an optimization model for resource allocation based on holistic criteria. [Article in preparation]

“Cutting libraries during a recession is like cutting hospitals during a plague.”

Eleanor Crumblehulme

This chapter documents early experiences of developing an optimal resource allocation model for distributing resources among different services and collection of a library. Specifically, this document addresses the problem of allocating funds for journal collections among divisions of an academic library. An optimization model for the problem is described with an objective of maximizing the usage of the journal collection over all library divisions subject to a single collection budget. To quantify usage, the study utilizes a combination of methodologies, that is, publishing patterns, citation analysis, and vendor-supplied statistics. An application of this model to an academic library in Belgium is presented as example of analysis.

Contributions of the first author

The first author's contributions are: introduction, theoretical background, related work, resource allocation model, and conclusions.

9.1 Introduction

Budget allocation is a core problem faced by all academic libraries independent of their size and of their funding mechanism, i.e., public or private (Arora & Klabjan, 2002). In fact, libraries in recent years have been threatened by tightened budget constraints (Sudarsan, 2006a; McKendrick, 2011; Guarria & Wang, 2011). This tendency stems from library services usually perceived as “free of charge” but in reality, not free of costs and strongly depends on institutional funding (Stouthuysen, Swiggers, Reheul, & Roodhooft, 2010). Moreover, although the migration of physical to digital environments has facilitated managing information and allowed the access to a number of digital journals and e-books; it has also contributed to escalating collection costs, as well as an increase in the complexity of budgeting models and resource allocation processes (Chan, 2008; Guarria, 2009; Poll, 2001). For instance, one of the problems with a subscription-based digital library collection and patron-driven acquisition is the variability of their yearly prices, which has rapidly risen in the last years (Allen Press, Inc., 2012). In fact, the most alarming trend in the academic library environment is the increase of information resource expenditures (Blake Gonzalez, 2011). Chan (2008) affirms that digital resource expenditures had a yearly average growth of 25%, while library budgets only had an average annual growth of 2.3%. These economic constraints result in a tremendous financial pressure for library directors, whom are required to shift budgeting and spending priorities (Blake Gonzalez, 2011). As a consequence, several decisions have been made, such as cutting collection budgets, eliminating budgets for travels or conferences, freezing salaries, and finding new ways to fund programs (Sudarsan 2006, McKendrick 2011).

In the current environment characterized by e-content revolution, technological advances, and ever-shrinking budgets, libraries are urged to identify efficient methods to allocate limited resources to collection and services that will provide the most benefit to users. Libraries more than ever must evolve and continue to demonstrate their relevance to institution management who face difficulties understanding the new roles, cost and value of good libraries (ACRL Research Planning and Review Committee, 2012, 2013). To do so, libraries have increased focus on assessment of outcomes over inputs and placed emphasis in demonstrating that these outcomes are having an impact on academic libraries and parent institutions. Libraries are also increasing their understanding of their users, collection and services, and related costs in order to justify resource requirements. Because of limited funding, library administrators are assessing the best ways to allocate their resources, how to redefine themselves, and reengineer their budget strategies.

In this context, budget allocation decisions for academic libraries become fairly important, but nevertheless a complex and difficult process, due to the diversity of constraints, data sources and formats required to be analyzed prior to decision-making, as well as the lack of efficient methods of integration (Kao, Chang, & Lin, 2003; Wu, 2003). Although many resource allocation approaches have been developed, most of them have mainly focused on the distribution of money for either physical or digital collections. There is also a lack of awareness on embracing different perspectives from heterogeneous stakeholders such as researchers, developers, administrators, librarians, and general users (Zhang, 2010). Therefore, scientific approaches on how to allocate limited resources among shifting collections and dynamic services become essential for libraries. The objective of this study is to document early experiences in developing an optimal resource allocation model for distributing funds among different services of a library system utilizing a holistic framework as data input. The remainder of this article will proceed as follows: the next section provides an

overview of the theoretical background, followed by a description of the methodology in the section after. The final section presents some concluding remarks and recommendations for future work.

9.2 Theoretical Background

In this section, the theoretical framework of the study is outlined with a summary of the literature review on the main aspects covered: holistic approach for data input, resource allocation and budgeting.

9.2.1 Holistic Approach for Data Input

David J. Ernst and Peter Segall (1995) state that institutions in challenging contexts and with limited budgets are called to develop strategic and well-coordinated budgeting plans by means of “holistic approaches”. The objective of a holistic approach is to help organizations define a set of measures that reflect their objectives and to assess their performance appropriately (Matthews, 2011). This holistic approach requires interconnecting all necessary components to evaluate the impact of limited resources in the whole institution, and then prioritize and optimize resource allocation in library services and collection. Many resource allocation approaches have been proposed; however, most of them mainly focus on economic allocation for either physical or digital collections separately. To the extent of our knowledge, the most complete approach to evaluate libraries from an holistic perspective is given by Scott Nicholson (2004). The author proposes a theoretical analysis framework to support libraries in gaining a more thorough and comprehensive understanding of their users and services for both digital and physical services. This theoretical analysis framework is based on a two-dimensional evaluation matrix, in which columns represent the topic: library system and use, and rows represent the perspective of the library system and users.

Due to the ease of understanding, completeness, and applicability to both physical and digital resources, Siguenza Guzman, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., and Cattrysse, D. (2015) adopted this framework as an input to an integrated decision support system. The main characteristics of each quadrant and the methodologies proposed by Nicholson and Siguenza-Guzman et al., shown in Figure 9.1 are described hereinafter. The first quadrant corresponds to *the internal perspective of the library system* that is to analyze the library performance, costs incurred, and resources consumed by library services. For that purpose, three costing system approaches are analyzed: traditional, activity-based, and time-driven activity-based, recommending the latter. The second quadrant evaluates *the external perspective of the library system*. Users’ perception about services quality is judged in this quadrant. To do so, one of the following methods is recommended: statistics gathering, suggestion boxes, Web usability testing, user interface usability, and satisfaction surveys. The third quadrant analyses *the external perspective of the library collection* that is to evaluate the impact of the current library collection on its users. A combination of the following three methods is proposed: citation analysis, vendor-supplied statistics, and citation databases. Eventually, the fourth quadrant evaluates *the internal perspective of the library collection*. The usage patterns followed to manipulate the library collection are analyzed. For this purpose, it is suggested to use log analysis methods such as transaction log analysis and deep log analysis. An overview of the proposed holistic evaluation framework is shown in Figure 9.1.

	<i>Library System</i>	<i>Library Collection</i>
<i>Internal Perspective (Library)</i>	1) Service Analysis Processes, cost, Time, Resources	4) Usage analysis Transaction log analysis, Deep log analysis
<i>External Perspective (Users)</i>	2) Quality Analysis Statistics gathering, Suggestion boxes, Usability testing, Satisfaction surveys	3) Collection Analysis Citation patterns, Publishing patterns, Journals downloaded, and Journal's impact factor

Figure 9.1: Methodologies proposed for the economic evaluation of libraries through a holistic approach (Siguenza-Guzman et al., 2015)

The resulting framework, as observed in Figure 9.1, requires retrieving and integrating information from various sources in order to be used in an adequate resource allocation process. By utilizing this simple but at the same time powerful framework for standardizing data input, this study ensures that the subsequent resource allocation efforts cover all library aspects in a comprehensive manner.

9.2.2 Resource allocation

Resource allocation, according to Barbara Blake Gonzalez (2011), is simply the most complex process of decision-making. In libraries, William B. Rouse (1975) argues that resource allocation can be performed in several levels. For instance, priorities within services or processes are defined on the lowest level, such as the number of librarians or computers for the reference services. At the intermediate level, the decisions are among the different services or processes. In this stage the concerns are about how to deal with the competition of resources, such as collection versus staff. The highest level relates to the competition between the library and other institutional departments. The objective in each stage of resource allocation is to assign funds in the most effective way in order to accomplish the objectives of the institution (Bookstein, 1974; Rouse, 1975).

The first step in any resource allocation process is defining performance indicators. The aim of a performance measurement is to assess the quality and effectiveness of services and collection provided by a library, as well as to assess the efficiency of resources allocated to such services and collection (Poll, 2001). Classic examples of library performance indicators are: availability of requested titles, collection turnover, and facility availability. Once a performance measure is defined for each of the services or activities, and a method for predicting changes in performance due to changes in resources is established, the next step is to consider optimizing the overall performance. For this purpose, an optimization criterion is required. To this end, it is necessary to involve the manager's decision-making criterion. This criterion will measure the different performance indicators based on the manager preference. Finally, an optimal resource allocation model is necessary to be defined, which maximizes the decision-maker's relative preferences for the measured performance (Rouse, 1975).

Although many resource allocation models can be found in literature, resource allocation decisions at most academic libraries are made using ad hoc approaches rather than attempting to model

decision-making criteria (Arora & Klabjan, 2002). Generally, funds for library collection development are allocated based on historical spending patterns and in some cases based on their own library priorities. In this way, libraries in many occasions have failed to answer the question of why more resources are allocated to one department collection than another. This historical spending pattern favors certain traditional academic departments and knowledge areas and negatively affects other newer and emerging departments within institutions. Additional elements in economic decision-making include strategic planning and budgeting. In turn, Bowen (1971) argues that planning and budgeting should be a closely integrated process.

9.2.3 Budgeting

A budget is an indispensable tool for management when aligning resource allocation with an institution's priorities. Unfortunately, budgeting is always a complex process since it has to deal with limited resources and growing requirements (Linn, 2007; Wise & Perushek, 1996). A budgeting process, unlike resource allocation, is usually a "top down" approach, as this is mostly directed by economic conditions and institution priorities expressed by both institutional authorities and library administrators. Typical budgeting activities involve: planning, control, coordination, communication, and prioritization of resource allocation. Academic libraries in particular struggle to make budget decisions in a time of scarce resources, dealing with budgets that tend to decrease or in the better case remain constant. In addition, aspects like inflation, new information requirements, and increased cost of materials make the library budgeting process rather difficult (Chan, 2008; Sudarsan, 2006).

Librarians have been investigating the best approach to allocate funds to their collection for a long time now. As a result, many budgeting system approaches to allocate resources can be found in literature, as for example: incremental line-item, formula based, mathematical decision model based, zero-based, program based, performance-based, responsibility center based, block-incremental based, and initiative-based budgeting (Linn, 2007). The main characteristics of these models are summarized in Table 9.1. In academic libraries, these budgeting systems are often mixed, incorporating two or more of these budgeting strategies (Blake Gonzalez, 2011; Linn, 2007). This combination is applied, for example, because one method can be used externally when requesting for funds and a different method can be used when distributing those funds internally. When considering these various options it is wise to keep in mind Green and Monical's (1985) remark, *"There are probably as many different ways of allocating resources in institutions of higher education as there are presidents of these institutions"*.

9.3 Budgeting in university libraries

Many approaches to support library decision-making for budget allocation have been proposed in literature. The methods used to support allocation decisions mostly include: formula-based models (Lowry, 1992; Niemeyer, Lawson, Millen, & Slattery, 1993; Bourgeois, Cohen, Dix, & Natesan, 1998), statistic-based models (Wu, 2003), goal-programming-based paradigms (Wise & Perushek, 1996, 2000), integer programming models (Smith, 1981), linear programming models (Goyal, 1973; Gleeson & Ottensmann, 1994; Arora & Klabjan, 2002; Abu Bakar, Rahman, & Yusop, 2011), queuing networks (Rouse, 1975, 1976; Smith, 1981), and performance-based models (Sudarsan, 2006).

Table 9.1: Definition, advantages and disadvantages of the budget system approach (Linn, 2007; Blake Gonzalez, 2011)

Budget Approach	Definition	Advantages	Disadvantages
Incremental line-item budgeting	Increase and decrease the budget equal for all units on a percentile basis.	<ul style="list-style-type: none"> - Widely used. - Tends to create the least amount of conflict during the process. - Relatively easy to create and to allocate money. 	<ul style="list-style-type: none"> - Static: strategic changes cannot be made to the budget without breaking its incremental nature. - Poor system to be used during a period of change.
Formula budgeting	Allocation of resources based on estimations of costs and applying one or more mathematical formulas.	<ul style="list-style-type: none"> - Relatively easy for the director in order to predict the amount of money that will be allocated. 	<ul style="list-style-type: none"> - Its rigidity makes it unlikely to foster innovative practices or new programs. - Lacks the flexibility needed when changes are made to the organization's mission.
Mathematical decision model-based budgeting	Created to help college administrators allocate money by using complicated computer models	<ul style="list-style-type: none"> - Effective model to determine the resources required for various needs. 	<ul style="list-style-type: none"> - The time that is needed to be invested is excessively long.
Zero-based budgeting	Budget system aimed at aligning expenditures with real and emerging needs. Its main characteristic is that the budget has to be entirely justified each year.	<ul style="list-style-type: none"> - Helpful tool when deciding how to allocate the funds among the units - It points out those expenses that are no longer necessary. Possibility of creation of modified versions related to this approach. 	<ul style="list-style-type: none"> - Takes a great deal of time to do the work that this system requires. - More likely to be instituted during a time of fiscal retrenchment rather than growth.
Program-based budgeting	Creation of budgets for particular programs instead of an entire department. Funds are allocated towards goal achievement/outputs via programs.	<ul style="list-style-type: none"> - Relatively easy to determine which programs are the least cost effective. 	<ul style="list-style-type: none"> - Hard to evaluate the outputs. - Quality of the services provided is not considered. - One must not only determine the costs and benefits of the various program options but also their comparative importance.
Performance-based budgeting	Focuses on outcomes rather than outputs. Focuses on activities that produce results. Strives to allocate resources on anticipated results.	<ul style="list-style-type: none"> - Easy not only to track the cost of each library service but also to determine its efficiency. 	<ul style="list-style-type: none"> - Difficult to trace back to the individuals units that helped to bring the outputs. - There is not a common criterion regarding what the higher education outcomes should be.
Responsibility center-based budgeting	Based on the principle that each unit must generate income equal or greater than what funds are expended.	<ul style="list-style-type: none"> - It forces units to pay for everything they use and to be paid for what they supply. 	<ul style="list-style-type: none"> - It can hamper cross-disciplinary work since each unit is so independent. - There is the definite risk that costly redundancies within the greater institution may develop.
Block-incremental-budgeting	An alternative to responsibility center-based. The spending part of the budget is decentralized, while the central administration more tightly controls the income.	<ul style="list-style-type: none"> - Unit heads have the flexibility to shift spending to where they think it is needed most. 	<ul style="list-style-type: none"> - A unit's budget may grow incrementally but its various budget lines may not.
Initiative-based budgeting	An organized way of creating a pool of money for funding new initiatives	<ul style="list-style-type: none"> - Forces units to reevaluate their activities to make sure that all of them are still needed. 	<ul style="list-style-type: none"> - It cannot be used indefinitely. - It forces units to give back a certain percentage of their base budget.

The most popular approaches are the formula-based (Lowry, 1992; Niemeyer et al., 1993; Bourgeois et al., 1998). Lowry (1992) provides a matrix formula that allocates funds for monographs and serials based on disciplinary needs and publishing patterns. This formula is the result of a cooperative effort of four institutions in the United States. The matrix formula utilizes twelve variables to perform its calculations, including, among others, credit hours, library usage, publication index, and books and serials costs. Niemeyer et al. (1993) report an experiment with materials budgeting using a formula that uses factors as: average costs of books and periodicals, number of titles available, number of faculty, and number of credit hours generated and degrees awarded. Bourgeois et al. (1998) improve an old formula-approach utilized for budget departmental allocation at the Southwest Texas State University Library by considering several factors, such as semester credit hours, number of students, degree programs and courses, and library usage rate. In these formula-based approaches, the allocation is determined proportionally based on selected factors that normally include library usage, collection costs, and academic features.

There have been algorithms based on programming techniques as well. Wise and Perushek (1996, 2000) utilize goal programming to allocate academic library's acquisitions funds based on the achievement of the library's goals and objectives. To this end, the model defines a set of goals, and then assigns a priority and a weight to each goal. The authors report a drawback of this model on its ability to deal only with a limited number of goals. Smith (1981) utilizes queuing networks, mixed integer programming, and expected utility theory to model the library building programming problem. The model utilizes a sample case study in the Champaign Public Library to allocate resources such as space, equipment, and staff to the card catalog and reference/information center of the library. The most popular programming-based approach is the linear system. Goyal (1973) describes the linear programming approach as a solution to the problem of allocation of funds to different university departments for purchasing books and journals. The model is based on a measure of the social benefits due to the allocated funds in a department, e.g., importance to the university and society and importance based on the size of the department. Gleeson and Ottensmann (1994) report a decision-support system that among its modules includes one for budget allocation. This module solves a continuous knapsack problem to calculate trial budgets. Librarians iteratively evaluate the trial budgets until the allocation is deemed satisfactory. Gleeson and Ottensmann's model, similar to the approach of Arora and Klabjan, is usage-based and utilizes forecasting to estimate monographs use. Arora and Klabjan (2002) address the problem of allocating funds for acquisition of periodicals among interrelated units at the Urbana-Champaign Library. Ho et al. (2010) formulate a linear programming problem with regard to materials budget allocation by using discrete particle swarm optimization to acquire optimal or near optimal solutions. The objective of the model is to maximize the average preferences of materials selection subjected to the constraints of material costs and required amounts in specified categories. Abu Bakar et al. (2011) suggest a budget allocation model that considers a balance between continuing commitment and new initiatives, between resources to support undergraduate learning and resources to support graduate work and research, and between subject disciplines. The approach's objective is to maximize the purchases subjected to budget allocated for books and journals.

Other approaches for allocation decisions include queuing networks and performance-based models. Rouse (1975) develops a procedure for optimal allocation of resources among processes of the Library Wessel at Tufts University by means of queuing theory. The model maximizes the

expected value of the decision makers' utility. Furthermore, Rouse (1976) utilizes a hypothetical network situation to characterize the performance of an interlibrary loan network by means of the probability of satisfying a request, delay in satisfying a request, total and unit costs, and processing load on each network number. Sudarsan (2006) develops a performance-based allocation model for university libraries in India. The model has two components, a base and an incremental component; the variables in the model are based on the number of students.

9.4 Towards a resource allocation model

This section begins to develop a metric to evaluate performance in a holistic manner and then to develop the corresponding model to optimize this metric for allocating resources in an academic library. More specifically, an allocation model is described that addresses the problem of allocating funds for journal collection among different divisions of an academic library in Belgium.

9.4.1 Case Study

The academic library under study is considered as one of the largest and most modern libraries into the areas of science and engineering in Belgium. It has a collection of one million books and reference works and additionally offers electronic and multimedia facilities. This academic library provides its services to the faculties of science, engineering, bioscience engineering, kinesiology, and rehabilitation sciences. The main library collection consists of two sections: a core and a research collection, organized in six clusters or divisions. To improve cost efficiency and effectiveness, this library has been forced to implement new strategies to deliver its services, such as the use of new technologies, improving access to e-journals and databases, automation of repetitive processes, and deployment of new digital and physical services. However, library budget cuts urge library management to keep improving its understanding and selection of the information collected for budget decision making. In order to maximize access to research information, this library currently tries to ensure access to the relevant top 20 journals per knowledge area¹⁷. Although, this current approach is advantageous because in some ways "facilitates" the allocation decision, and ensures access to the most significant recent thinking in the field; this approach is quite limited in terms of finances, since it can give the impression that price does not matter. An additional limitation of this approach is that it is not personalized to the local information needs.

9.4.2 Usage analysis

As a first approach, the resource allocation model assumes that the academic library has a single budget for journal collection that has to be allocated to several knowledge areas. The objective of the model is to distribute funds across different library divisions in a way that the usage of journal collection is maximized. To this end, the present study follows the methodology employed by Anish Arora and Diego Klabjan (2002) to allocate funds for acquisition of periodicals among several units of an academic library. This approach presents several challenges. The first challenge to be addressed is how to quantify usage. According to Siguenza-Guzman et al. (2015), usage in a digital

¹⁷ <https://bib.kuleuven.be/2bergen/cba/2bergen/missie-visie-en-strategische-doelstellingen-voor-de-bibliotheek-cba-en-biomedische-bibliotheek>

environment can be measured using transaction and deep log analysis. Unfortunately, to date, no studies have assessed usage behavior in this academic library (Siguenza-Guzman et al., 2013). Consequently, alternative measures of usage are taken into consideration, such as to evaluate the third quadrant of the holistic matrix, i.e., the usefulness of the library collection. To do so, a combination of citation analysis, vendor-supplied statistics, and citation databases, such as PubMed, Scopus, Web of Science, and Scopus are used.

Data used by this model is obtained from an internal project that combines these three methodologies (Siguenza-Guzman et al., 2013). This project analyzes more than 1,200 PhD theses submitted over a six-year period (2005-2010). These theses correspond to research conducted in 13 departments of Science, Engineering and Agriculture at KU Leuven. To this end, the study first collected in a database all *references cited* in each PhD thesis. In parallel, a second database was created gathering information about the *publishing patterns* of PhD researchers. This second database determined the most attractive journals where departments choose to publish, as well as verifying whether these journals correlated with the citations used as reference. A third database was used to collect the *vendor-supplied statistics* of all journals downloaded during the period 2005-2010. These electronic journal usage data were received from COUNTER-compliant publishers as part of the subscription contract. The Counting Online Usage of NeTworked Electronic Resources (COUNTER) standards are an internationally accepted initiative that facilitates the recording and exchange of online usage data in a consistent, credible, and compatible manner (COUNTER, 2013). This third database verifies the correlation among the citations patterns, publishing patterns, and journals downloaded. The information collected in these three databases is also used to test an additional correlation with the *5-year Impact Factor (IF)* of the Journal Citation Reports (JCR) produced by Thomson ISI Web of Knowledge. This commercial database contains information on journal rankings, citation impacts, and the reputation of academic journals.

As an initial attempt, a percentage of increase over the number of citations/references, publishing patterns, and vendor-supplied statistics is utilized as the basis to calculate the future usage for the forthcoming year. After consulting library management and other staff members, some initial assumptions were made: 1) The use of electronic resources has an average annual increase of 5% (on and off campus). Although this assumption can reduce the dynamic behavior, further improvements to the model include forecasting usage values. 2) The number of publications per journal is considered more important than the number of citations and vendor-supplied statistics; therefore, the relation among these values is estimated as: 50:30:20% respectively. Further improvements on this approach would include sensitivity analysis to show the impact of these initial assumptions. The usage value per journal is calculated as follows:

$$u = 0.3Uc + 0.5Up + 0.2Uv$$

where,

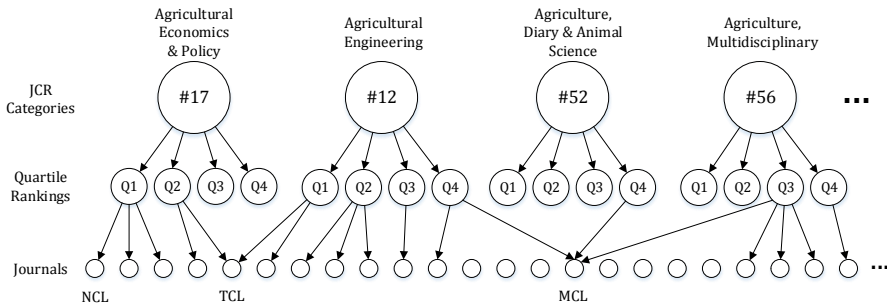
u	usage value per journal
Uc	usage value based on citation analysis
Up	usage value based on publishing patterns
Uv	usage value based on vendor-supplied statistics

A general drawback in this approach is that the number of journals in the sample is dynamic, i.e., increases or varies over the years. Therefore, following the recommendation of Arora and Klabjan, journal titles were first grouped based on subjects' areas or categories, and then these categories

were used in the objective function to make decisions on the number of journals to be purchased per category. To categorize journals, the JCR database was used, in which journals are classified based on subject categories such as Mathematics, Agronomy, Biology, and so on.

The second challenge is to combine the data for citations, publications, and vendor-supplied statistics as a measure of usage. According to Arora and Klabjan, a general strategy is to assign weights to each number and use the combined data in the model. Thus, each subject category can have its own weight. However, the same authors realized that assigning weights is extremely difficult due to the large number of these subject categories. In order to overcome this difficulty, we propose to split each category into four subcategories based on the quartile rankings of the JCR database, that is, Q1, Q2, Q3, and Q4. Quartile rankings are derived for each journal in each of its subject categories according to which quartile of the IF distribution the journal occupies for that subject category. Q1 indicates the top 25% of the IF distribution, Q2 for middle-high position, i.e., between top 50% and top 25%, Q3 middle-low position (top 75% to top 50%), and Q4 the lowest position, namely, bottom 25% of the IF distribution. Therefore, journals dominating major research fields are categorized as Q1 and Q2 journals, while the more general scoping journals are classified as Q3 and Q4 journals.

An additional challenge that arises when categorizing journals is that the JCR classification assigns journals to more than one category; therefore, a strategy to avoid double-counting is required. Aurora and Klabjan differentiate three types of behaviors: 1) *non-cross-listed*, when a journal is assigned to only one subcategory; 2) *twice-cross-listed*, a journal is assigned to exactly two subcategories; and 3) *more-cross-listed*, a journal is assigned to more than two subcategories. A graphical representation of this structure classification is depicted in Figure 9.2.



Note: NCL = Non-Cross-Listed; TCL = Twice-Cross-Listed; MCL = More-Cross-Listed

Figure 9.2: Structure classification used for determining the weights

Based on these three behaviors, usage values per journal of subcategory s are calculated as follows:

$$\hat{u}_i = w_i u_i \quad \forall i$$

$$\tilde{u}_{ij} = \frac{w_i u_i + w_j u_j}{2} \quad \forall i, j$$

$$\bar{u}_k = \frac{\sum_t w_t u_t}{|S_k|} \quad \forall k$$

Subject to

$$0 \leq w_i \leq 1 \quad \forall i$$

where,

i	subcategories where $i \in S$
j	subcategories where $j \in S$
k	subcategories where $k \in S_k$
S	set of subcategories
S_k	set of subcategories that the journal k is assigned to, where $S_k \subseteq S$
K	set of more-cross-listed journals
\hat{u}_i	usage value per non-cross-listed journal of subcategory s
\hat{u}_{ij}	mean usage value per twice-cross-listed journal of subcategory s
\bar{u}_k	mean usage value per more-cross-listed journal of subcategory s
w_i	weight assigned to journal quartiles $\{Q1, Q2, Q3, Q4\}$

9.4.3 Formulating the model

Figure 9.3 illustrates a graphical representation of the resource allocation problem. The notations used for this problem are the following:

- *Indices*

i	journals that are non-cross-listed where $i = 1$ to I_s
j	journals that are twice-cross-listed where $j = 1$ to J_s
k	journals that are more-cross-listed where $k = 1$ to K_s
s	subcategories where $s = 1$ to S

- *Decision variables*

x_i	non-cross-listed journals where $x_i = 1$ if the journal should be purchased, $x_i = 0$ if the journal should not be purchased for every journal i ; $i = 1$ to I_s (Binary)
y_j	twice-cross-listed journals where $y_j = 1$ if the journal should be purchased, $y_j = 0$ if the journal should not be purchased for every journal j ; $j = 1$ to J_s (Binary)
z_k	more-cross-listed journals where $z_k = 1$ if the journal should be purchased, $z_k = 0$ if the journal should not be purchased for every journal k ; $k = 1$ to K_s (Binary)

- *Parameters*

\hat{c}_i	price of a journal i ; $i = 1$ to I_s
\hat{c}_j	price of a journal j ; $j = 1$ to J_s
\hat{c}_k	price of a journal k ; $k = 1$ to K_s
n_s	maximum number of journals available in subcategory s
p_s	minimum number of journals available in subcategory s
B	budget for the next fiscal year
S	set of subcategories
I_s	set of non-cross-listed journals in subcategory s
J_s	set of twice-cross-listed journals in subcategory s

K_s

set of more-cross-listed journals in subcategory s

		Subcategories (S)											
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	...	
Non-cross-listed (I)	X1	1											Number of Journals (n)
	X2	1											
	X3	1											
	X4		1										
	X5		1										
	...												
Twice-cross-listed (J)	Y1	1		1									
	Y2	1	1										
	Y3	1			1								
	...												
More-cross-listed (K)	Z1	1		1	1			1					
	Z2	1	1			1							
	...												

Library Division (I1)

Library Division (I2)

Figure 9.3: Exemplification of the resource allocation problem

The resource allocation problem can be modeled as follows:

$$\max \sum_i \hat{u}_i x_i + \sum_j \hat{u}_j y_j + \sum_k \bar{u}_k z_k \quad (1)$$

Subject to

$$\sum_i \hat{c}_i x_i + \sum_j \bar{c}_j y_j + \sum_k \bar{c}_k z_k \leq B \quad (2)$$

$$p_s \leq \sum_{i \in I_s} x_i + \sum_{j \in J_s} y_j + \sum_{k \in K_s} z_k \leq n_s \quad \forall s \quad (3)$$

$$\sum_i x_i \leq n_s \quad \forall s \quad (4)$$

$$\sum_j y_j \leq n_s \quad \forall s \quad (5)$$

$$\sum_k z_k \leq n_s \quad \forall s \quad (6)$$

$$x_i \in \{0,1\} \quad \forall i \quad (7)$$

$$y_j \in \{0,1\} \quad \forall j \quad (8)$$

$$z_k \in \{0,1\} \quad \forall k \quad (9)$$

The objective function (1) maximizes the usage of journal collection in the future fiscal period subject to: the constraint that (2) funds are not allocated more than the available budget; (3) a minimum number of journals in each subcategory is billed, and no more journals than the maximum available in each subcategory are allocated; (4), (5) and (6) no more non-cross-listed, twice-cross-listed, and more-cross-listed journals respectively than the maximum available in each

subcategory are allocated; and (7), (8) and (9) decision variables for non-cross-listed, twice-cross-listed, and more-cross-listed journals respectively only receive 0 or 1 values.

Finally, an additional constraint (10) ensures that funds are allocated to a library division given lower and upper bounds.

$$\underline{b}_l \leq \sum_{i \in A_l} \hat{c}_i x_i + \sum_{j \in B_l} \tilde{c}_j y_j + \sum_{k \in C_l} \bar{c}_k z_k \leq \bar{b}_l \quad \forall l \quad (10)$$

- *Indices*

- i non-cross-listed journals that are (or would be) in library division l
where $i = 1$ to A_l
- j twice-cross-listed journals that are (or would be) in library division l
where $j = 1$ to B_l
- k more-cross-listed journals that are (or would be) in library division l
where $k = 1$ to C_l
- l library divisions where $l = 1$ to L

- *Parameters*

- \underline{b}_l minimum amount of budget allowed to be allocated to library division l
- \bar{b}_l maximum amount of budget allowed to be allocated to library division l
- L set of library divisions
- A_l set of non-cross-listed journals in library division l
- B_l set of twice-cross-listed journals in library division l
- C_l set of more-cross-listed journals in library division l

9.5 Conclusions

This article describes a preliminary model to distribute resources among different processes of a library system utilizing a holistic framework as data input. This preliminary approach has faced several challenges. The first main challenge was how to quantify usage. To do so, a combination of methodologies coming from the third quadrant of the holistic matrix is recommended, that is, citation analysis, vendor-supplied statistics, and citation databases. As an initial attempt, a percentage of increase over the usage values is utilized as the basis to calculate the future usage. The second challenge was to combine the data coming from citations, publications, and vendor-supplied statistics as a measure of usage. Ideally, a general strategy would be to assign weights to each number and use the combined data in the model; thus, each subject area can have its own weight. However, assigning weights to each subject area is extremely difficult due to the large number of these areas. To overcome these two challenges, journal titles were first grouped based on subject areas or categories, and then these categories are divided into four subcategories based on the quartile rankings of the JCR database. These subcategories are used in the objective function to make decisions on the number of journals to be purchased per subcategory. Finally, an additional challenge that arose when categorizing journals was that using the JCR classification, journals can be assigned to more than one category; therefore, a strategy to avoid double-counting was required. In addition, this study only discusses how a library should manage the budget for journals. The expenses incurred on books, salaries, maintenance, and other indirect costs were not considered in this initial model. Although this is just an initial approach toward a complete solution, we may conclude that this optimization technique seems feasible, with potential benefits for library managers. In addition, this stage of the system allows researchers to identify opportunities for

future studies and applications, such as incorporating log analysis results and interlibrary loan requests. The implementation of the complete solution based on the holistic approach is definitely a must for future research.

References

- Abu Bakar, E. M. N. B. E., Rahman, S. A., & Yusop, N. M. (2011). Modelling of Budget Allocation for University Library. *Journal of Statistical Modeling and Analytics*, 2(2), 1–8.
- Arora, A., & Klabjan, D. (2002). A model for budget allocation in multi-unit libraries. *Library Collections, Acquisitions, and Technical Services*, 26(4), 423–438. <http://doi.org/10.1080/14649055.2002.10765879>
- Blake Gonzalez, B. (2011). Resource Allocation Strategies In Doctoral/Research University (extensive) Libraries. THE GEORGE WASHINGTON UNIVERSITY. Retrieved from <http://gradworks.umi.com/34/33/3433988.html>
- Bookstein, A. (1974). Allocation of resources in an information system. *Journal of the American Society for Information Science*, 25(1), 52–58. <http://doi.org/10.1002/asi.4630250107>
- Bourgeois, E., Cohen, P., Dix, J., & Natesan, C. (1998). Faculty-Determined Allocation Formula at Southwest Texas State University. *Collection Management*, 23(1-2), 113–123. http://doi.org/10.1300/J105v23n01_05
- Bowen, W. G. (1971). The Role of the Business Officer in Managing Educational Resources. The Management Challenge: Now and Tomorrow. Managing Educational Programs. NACUBO Professional File, 2(3). Retrieved from <http://eric.ed.gov/?id=ED113993>
- Chan, G. R. Y. C. (2008). Aligning collections budget with program priorities: A modified zero-based approach. *Library Collections, Acquisitions, and Technical Services*, 32(1), 46–52. <http://doi.org/10.1016/j.lcats.2008.06.001>
- COUNTER. (2013, February). About COUNTER. Retrieved from <http://www.projectcounter.org/>
- Ernst, D. J., & Segall, P. (1995). Information Resources and Institutional Effectiveness: The Need for a Holistic Approach to Planning and Budgeting. *Cause/Effect*, 18(1), 11–13.
- Gleeson, M. E., & Ottensmann, J. R. (1994). A Decision Support System for Acquisitions Budgeting in Public Libraries. *Interfaces*, 24(5), 107–117.
- Goyal, S. K. (1973). Allocation of Library Funds to Different Departments of a University--An Operational Research Approach. *College and Research Libraries*.
- Green, J. L., & Monical, D. G. (1985). Resource allocation in a decentralized environment. *New Directions for Higher Education*, 1985(52), 47–63. <http://doi.org/10.1002/he.36919855206>
- Ho, T.-F., Shyu, S. J., Wu, Y.-L., & Lin, B. M. T. (2010). Discrete Particle Swarm Optimization for Materials Budget Allocation in Academic Libraries. In *Proceedings of the 2010 13th IEEE International Conference on Computational Science and Engineering* (pp. 196–203). Washington, DC, USA: IEEE Computer Society. <http://doi.org/10.1109/CSE.2010.33>
- Kao, S. C., Chang, H. C., & Lin, C. H. (2003). Decision support for the academic library acquisition budget allocation via circulation database mining. *Information Processing & Management*, 39(1), 133–147. [http://doi.org/10.1016/S0306-4573\(02\)00019-5](http://doi.org/10.1016/S0306-4573(02)00019-5)
- Linn, M. (2007). Budget systems used in allocating resources to libraries. *Bottom Line: Managing Library Finances*, The, 20(1), 20–29. <http://doi.org/10.1108/08880450710747425>
- Lowry, C. B. (1992). Reconciling Pragmatism, Equity, and Need in the Formula Allocation of Book and Serial Funds. *College & Research Libraries*, 53(2), 121–138. http://doi.org/10.5860/crl_53_02_121
- Matthews, J. R. (2011). Assessing Organizational Effectiveness: The Role of Performance Measures. *The Library Quarterly*, 81(1), 83–110. <http://doi.org/10.1086/657447>
- Nicholson, S. (2004). A conceptual framework for the holistic measurement and cumulative evaluation of library services. *Journal of Documentation*, 60(2), 164–182. <http://doi.org/10.1108/00220410410522043>
- Niemeyer, M., Lawson, L., Millen, G., & Slattery, C. (1993). Balancing Act for Library Materials Budgets. *Technical Services Quarterly*, 11(1), 43–60. http://doi.org/10.1300/J124v11n01_05

- Poll, R. (2001). Performance measures for library networked services and resources. *Electronic Library*, The, 19(5), 307–315. <http://doi.org/10.1108/02640470110408887>
- Rouse, W. B. (1975). Optimal resource allocation in library systems. *Journal of the American Society for Information Science*, 26(3), 157–165. <http://doi.org/10.1002/asi.4630260304>
- Rouse, W. B. (1976). A library network model. *Journal of the American Society for Information Science*, 27(2), 88–99. <http://doi.org/10.1002/asi.4630270204>
- Siguenza-Guzman, L., Holans, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2013). Towards a holistic analysis tool to support decision-making in libraries. In *Proceedings of the IATUL Conferences* (p. Paper 29). Cape Town, South Africa: Purdue e-Pubs. Retrieved from <http://docs.lib.purdue.edu/iatul/2013/papers/29/>
- Siguenza-Guzman, L., Van Den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2015). A holistic approach to supporting academic libraries in resource allocation processes. *The Library Quarterly: Information, Community, Policy*, 85(3).
- Smith, J. M. (1981). The use of queuing networks and mixed integer programming to allocate resources optimally within a library layout. *Journal of the American Society for Information Science*, 32(1), 33–42. <http://doi.org/10.1002/asi.4630320105>
- Sudarsan, P. K. (2006). A resource allocation model for university libraries in India. *Bottom Line: Managing Library Finances*, The, 19(3), 103–110. <http://doi.org/10.1108/08880450610682527>
- Wise, K., & Perushek, D. E. (1996). Linear goal programming for academic library acquisitions allocations. *Library Acquisitions: Practice & Theory*, 20(3), 311–327. [http://doi.org/10.1016/0364-6408\(96\)00065-8](http://doi.org/10.1016/0364-6408(96)00065-8)
- Wise, K., & Perushek, D. E. (2000). Goal Programming as a Solution Technique for the Acquisitions Allocation Problem. *Library & Information Science Research*, 22(2), 165–183. [http://doi.org/10.1016/S0740-8188\(99\)00052-3](http://doi.org/10.1016/S0740-8188(99)00052-3)
- Wu, C.-H. (2003). Data mining applied to material acquisition budget allocation for libraries: design and development. *Expert Systems with Applications*, 25(3), 401–411. [http://doi.org/10.1016/S0957-4174\(03\)00065-4](http://doi.org/10.1016/S0957-4174(03)00065-4)

PART V

Conclusions

Chapter 10: Conclusions

“The secret of change is to focus all of your energy, not on fighting the old, but on building the new.”

Socrates

10.1 Introduction

Libraries since their inception 4000 years ago have been in a process of constant change. Although for centuries, momentous changes were in slow motion, in the last decades, academic libraries, specifically, have been continuously striving to adapt their services to the ever-changing user needs of students and academic staff. Library users have changed their information-seeking behaviors due to rapid technological advances and the astonishing e-content revolution; the growing presence of e-books and the proliferation of tablets and mobile devices have transformed the manner how information is disseminated and consumed. Furthermore, e-services like remote access to digital information have meant that many students and scholars misunderstand what libraries do for them and not necessarily associate the library with providing information resources. Consequently, libraries, recognizing that their intrinsic information provider role in the current evolving information environment is becoming less and less visible, have responded to these challenges and technological developments by rethinking and repurposing what libraries are and what libraries do for their users. Moreover, although the migration of physical to digital environments has facilitated managing information and allowed the access to a number of digital journals and e-books; it has also contributed to escalating collection costs, as well as increasing the complexity of budgeting models and resource allocation processes.

This panorama evidences the stressing time for libraries to be more innovative in providing, justifying, and evaluating the efficiency and effectiveness of their services and collection. Libraries, more than ever, must evolve and continue to demonstrate their relevance to the academic management, who faces difficulties understanding new roles, cost, and value of good libraries. To do so, libraries have increased focus on assessment of outcomes over inputs, and placed emphasis in demonstrating that these outcomes have an impact on academic libraries and parent institutions. Libraries require an increasing understanding of their users, collection and services, and related costs in order to justify resource requirements. The first research question (RQ) presented in this dissertation addresses the recompilation of relevant data from libraries, in a systematic manner, to support management in making optimal budgeting and resources allocation decisions.

10.2 Summary of significant findings with respect to the formulated research propositions

This section presents a summary and analysis of the four research questions raised in the introductory Chapter 1. Additional components of this section include lessons learned drawn from

deploying the different blocks of the decision-support system as well as future directions in each research question.

10.2.1 Data Source

RQ1: How to collect data in a structured manner, covering the key aspects of a library and, at the same time, facilitating the understanding and replication of the data collection process.

One of the most challenging aspects for the implementation of an integrated decision-support system is its data collection. To tackle this challenge, a four-pronged theoretical framework is used, in which the library system and collection are analyzed from the perspective of users and internal stakeholders. Based on this theoretical framework, in Chapter 2, a holistic structure and the required toolset to holistically assess libraries are proposed as a baseline to collect and organize the data from an economical point of view. An overview of the proposed holistic evaluation framework is shown in Figure 10.1. The proposed approach aims to provide an integrated solution to assist library managers to make economic decisions based on a perspective of the library situation that is as realistic as possible. This theoretical matrix is used as a reference to analyze the library collection and services from internal and external perspectives. Furthermore, several methods and appropriate measurement tools have been evaluated and proposed for an integrated decision-making process. Library managers can select one or more instruments in every quadrant based on the current availability or even decide to include other measurements in the model.

	<i>Library System</i>	<i>Library Collection</i>
<i>Internal Perspective (Library)</i>	Service Analysis Processes cost, Time, Resources	Usage Analysis Transaction log analysis, deep log analysis
<i>External Perspective (Users)</i>	Quality Analysis Statistics gathering, Suggestion boxes, Usability testing, Satisfaction surveys	Collection Analysis Citation patterns, Publishing patterns, Journals downloaded, and journal's impact factor

Figure 10.1: Methodologies proposed for the economic evaluation of libraries through a holistic approach

An example reporting the preliminary experiences, benefits and challenges of organizing and collecting library data based on the four quadrants of this holistic approach in the Arenberg Campus Library (CBA) of KU Leuven in Belgium is presented in Appendix A. In this example, the first quadrant documents the experience of analyzing the processes and services of several library functions. Because this aspect deserves much-more attention, a deep analysis of this quadrant is provided in the following research question. The second quadrant describes how CBA dealt with several challenges to deploy a quality analysis, such as: issues of language definition and great variety of population; granularity issues, as no specific results for branch libraries and disciplines are provided by standard reports of the quality system; and low participation rates due to the perception that the survey was very long. The third quadrant in this library is approached through a combination of three types of collection analysis: citation analysis, publishing patterns, and vendor-supplied statistics. The challenges faced during the implementation of this analysis were: 1) the amount of time necessary to analyze a single document; 2) lack of standards for journal's abbreviations; and 3) the need of dedicated software to collect the large amount of information, as well as to evaluate the results. Finally, the fourth quadrant that measures users' interaction with the system presented several challenges that have yet to be broadly resolved, such as issues of privacy and confidentiality. All in all, this initial experience shows that the proposed holistic model and toolset constitute a simple and powerful structure for grouping library information. Nevertheless, important considerations need to be borne in mind, mainly the time required to implement the complete approach and the need for dedicated systems to automate data collection across the different quadrants of the holistic matrix.

In the case study of the Regional Documentation Centre “Juan Bautista Vazquez” (CDRJBV) in Ecuador, challenges presented to implement the full holistic framework were slightly different. For instance, similar to the previous case, no prior studies have analyzed service performance as required in the first quadrant. Unlike the standardized on-line survey that CBA utilizes for quality analysis, CDRJBV developed its own quality survey system. Although this software can be considered a good first attempt; the registration to the standardized web survey is highly recommended. Data to feed the model in the third quadrant were initially limited to the use of vendor-supplied statistics; the citation database Scopus was recently made available. Eventually, in contrast to CBA, data for the fourth quadrant were feasible to collect due to the few existing regulations with respect to privacy concerns. On the whole, this second implementation experience reinforced the need for using additional case studies in order to validate the model and have a broader perspective of the libraries’ situation.

During the data collection experience, it was observed that libraries are used to collect statistics and data, extensive enough to fill all the quadrants of the proposed holistic assessment structure. However, one of the key elements to support economical decisions is the cost allocation of the different services and activities performed within libraries. For some aspects, like library collection, the cost is normally the same no matter how often the collection is accessed because of a fixed subscription and purchasing cost; but for other library services, it presents a great challenge. Libraries in general are unfamiliar with performing formal costing analyses to their services and processes. Therefore, RQ2 addresses the issue of implementing cost analyses in a formal and accurate manner, while keeping the burden to a minimum.

10.2.2 Costing analysis

RQ2: How to calculate the cost of library services based on a formal costing analysis in a way that can be widely and effectively applied, while minimizing the required resources.

Following a comprehensive literature review on cost analysis (Chapter 3), in which thirty-six case studies were analyzed and classified along application themes such as logistics, manufacturing, services, health, hospitality, and other nonprofit services, it is concluded that TDABC is highly recommended for evaluating repetitive activities. Comparing TDABC to the traditional ABC costing, TDABC offers several advantages, even if it does not dramatically simplify specific processes. However, the analyzed research is less clear about the advantages of TDABC for non-routine tasks. Technologies such as RFID, bar codes, or existing information from time sheets may provide the necessary data required in these cases. It is worth noting that the studies on the implementation, as well as the criticisms on TDABC, in most of the cases, are written by its creators and not by independent researchers. This can certainly bias the evaluation of the TDABC methodology. Therefore, the need of more research by means of operational case studies in specific areas, such as public services and in activities that follow unstructured and non-systematic sequences, has been identified. Consequently, several case studies of the implementation of TDABC in academic libraries were performed in the course of this research.

Case studies conducted on loan and return processes (Chapter 4) as well as on cataloging processes (Appendix B) of the Arenberg Campus Library of KU Leuven illustrate, through six simple steps, how TDABC can be used in carrying out a cost analysis in a simple, easy-to-understand and accurate manner. Several important insights have also emerged from the case study. The first important insight is that although the amount of time required to collect and document the duration of activities and activity flows is relatively long compared to traditional costing systems, the insights gained from the analysis are more compelling and robust. The duration of activities was gathered by direct observation since the most accurate data were collected when librarians physically performed the tasks. Although, initially, this process is more time consuming, nonetheless, the final model considers real and detailed values regarding the library activities. Therefore, a trade-off between measurement time and accuracy must be reached.

A second important insight is that software tools and the ease of presenting results help to decrease implementation time and allow for better communication and validation. MS Office suite programs,

such as Visio and Excel, were integrated to store, analyze, and create graphical representations of activity flows. As a consequence of this clear graphical representation, librarians were easily able to understand the sequences and their responsibilities in each process; and consequently, this allowed us to validate the collected information straightforwardly. However, a dedicated software tool to perform TDABC analysis was strongly recommended in order to keep the flows updated and consequently to facilitate long-term maintenance. Following this recommendation, a web-based software tool TD-ABC-D for TDABC analysis in libraries processes was deployed as described in Appendix C. Finally, a third important insight is that the involvement and commitment of the library staff are critical to the data collection in increasing the acceptance of the model. Therefore, motivation and an explanation of the measurement purpose are fundamental to achieving the desired commitment needed from the staff. In the case of a large library, this requirement is even more critical since the number of employees gives rise to different types of opinions and attitudes regarding the process. For instance, for some librarians, the disclosure of information on their salaries was deemed to be a very sensitive question. Therefore, the case study shows that TDABC is applicable to large libraries as well but that the involvement of library staff is crucial.

Despite the TDABC benefits arising from implementing this model, a number of challenges were found during the TDABC implementation. For instance: 1) Time: data collection on the duration of activities took significant time, as the measuring was gathered by direct observation. Moreover, documenting the activity flows required considerable time. Two rounds of interviews were conducted with library managers and staff in order to identify the activities, resources and responsible. 2) Feeling controlled: some staff members felt uncomfortable being observed while working. This discomfort caused some resistance and consequently delayed the data collection. A right communication as well as the involvement and commitment of the managers and staff can increase the level of acceptance. In addition, library managers and TDABC team should explain the purpose of measurement, importance of the model, activities to perform, and implications of the results. They should clearly state that the activities and profiles are measured, not the names of individuals.

In summary, although at first glance, TDABC may seem more difficult to implement and requires more intensive data collection compared to a traditional costing system, the presented investigations show that TDABC in practice is simple and easy to understand when the six steps identified by Everaert and colleagues are followed. Furthermore, the potential benefits accruing from the TDABC implementation, such as the accuracy to calculate the costs of library services, the possibility of performing benchmarking analysis, disaggregating values per activity, and justifying decisions and choices, validates the effort required to collect the data. An interesting avenue for future research resulting from the case studies implementation is to perform process benchmarking utilizing TDABC. Benchmarking library processes provides real evidence that additional resources, technological and logistic changes, or support for infrastructure are needed. Internal benchmarking can be potentially used to better manage local processes by measuring and tracking their changes, to justify allocation and prioritization decisions, and to enable assessment activities.

As a consequence of the previous recommendation, a benchmarking study was conducted for two Belgian libraries (Chapter 5). One of the most significant gains from this analysis is to make evident that TDABC makes available information about the cost of providing services, and disaggregates their corresponding causes. TDABC not only provides library managers with holistic information to make sound decisions but also with enough tools and strategic information to agilely identify improvement opportunities. In the Time-Driven benchmarking, processes of library 1 and library 2 are compared in time and cost in order to highlight the best practices of both libraries. In the absence of TDABC analysis, the manager of library 1 can wrongly assume that in overall macro results, this library outperforms library 2 in all aspects and that nothing needs to be changed in its processes. However, the performed benchmarking illustrates how both libraries have to learn from each other if wheels are not to be reinvented on both sides. Thus, mutually beneficial ways of improving library performance can be found through this type of comparison. Library 1, for instance, should focus on improving the scanning equipment for the ILL services and eliminating the non-value-added steps coming from old ("legacy") procedures, such as printing and storing request forms. On the contrary, library 2 should focus on facilitating data entry into the LIS,

relocating the closed stack collection, and delegating more responsibilities to low-wage employees like library assistants or students. Time-Driven benchmarking encourages rethinking roles, rules, and activities across the library workflow without spending time on problems that have already been solved by exchanging knowhow among libraries. It helps to rethink how time is spent within library processes, improve or streamline processes, reduce variability, and standardize workflows. Despite positive implications and results, some limitations of this benchmarking deserve consideration. First, although process improvements can be identified throughout comparative analysis, in some cases, certain aspects such as physical infrastructure and transportation distances cannot be easily changed or adapted. Second, even though both libraries provide comparable services and have similar levels of automation, each library may have different priorities. For instance, library 1 may emphasize quality in original cataloging, whereas library 2 focuses on fast copy cataloging or any other digital service.

Once the appropriate methodological tool for cost analysis is identified and tested through various case studies, the data collection framework is fully defined. With a complete framework to collect data, there comes the challenge to properly integrate and store such data. This integration poses a great challenge that needs to be addressed since different data sources normally use dissimilar formats and access methods. This issue is tackled in RQ3.

10.2.3 Data Storage

RQ3: What architecture is adequate to store the data collected through the holistic approach from different sources and formats that enables to analyze and maintain big quantities of data?

Today, libraries have an excessive amount of data to be processed that complicates their management and consultation for decision-making. A data warehouse (DW), being a platform that integrates strategic data, supports this decision-making process by allowing to have a global picture of the organization. Therefore, a data warehousing approach is proposed in the study to consolidate, filter, and process all the information extracted from many different systems and formats (Chapter 6). Based on the holistic approach proposed and the methodology and technologies selected to implement the DW, an integrated decision support system (iDSS) architecture is developed. Figure 10.2 shows its main elements. The proposed architecture allows the use of information, not only in traditional measures or for generating reports but also to enhance decision-making. For instance, information on the following four scenarios is accessible: 1) redistributing and prioritizing the allocation of resources assigned to a specific service; 2) gaining knowledge about users coming into the library and also users who are served by digital services; 3) awareness of the gaps and strengths in services and collections; and 4) the building of collections based on a library's holdings, users priorities, and technological tendencies. The advantage of this approach is that it increases effectiveness when more data are available. The more historical data are provided, the more accurate results will be obtained.

A case study to demonstrate the applicability of the holistic approach proposed to implement an iDSS based on DW was conducted at CDRJBV. The main contribution of this work is the analysis and design of an iDSS for a university library through the case study analyzed. The distinguishing feature of the proposed architecture is the emphasis on the use of a holistic conceptual matrix to select the corresponding data sources. By so doing, services and processes currently unavailable in the library such as book reservations, interlibrary loans, fines, and quality surveys could be considered for future inclusion in the model in order to increase scalability and also to meet the constraints of limited resource environments as in this case study. Therefore, the decision of using the holistic approach for data input implied integrating data from multiple and heterogeneous sources from the library, university, consortiums, and suppliers, all of which use dissimilar formats and access methods including both structured and unstructured data. Consequently, an adequate selection of methodology and technological tools for constructing the DW is necessary to ensure the data warehousing success. Important to note is that, thanks to the use of the Hefesto methodology at early deployment time, library managers and stakeholders were able to realize the potential of implementing an iDSS solution in order to make tactical decisions about the optimal use and leverage of their resources and services. Hefesto is an easy methodology that provided necessary

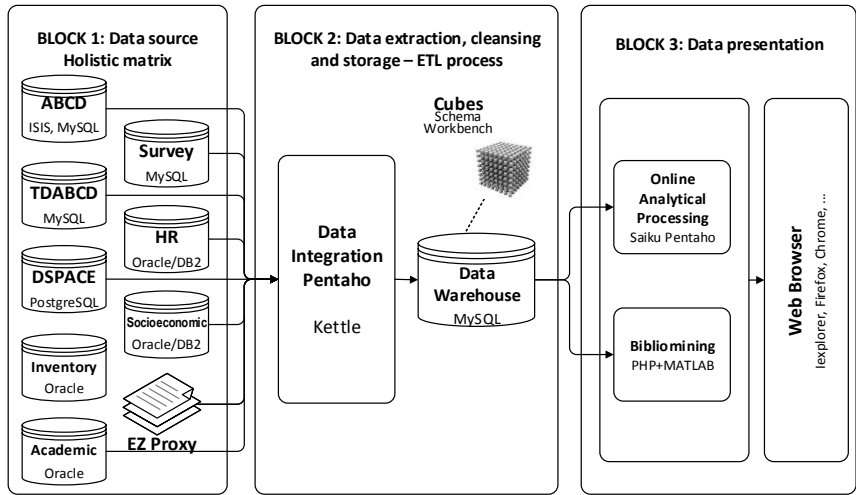


Figure 10.2: iDSS architecture of the CDRJBV

steps for an efficient DW implementation. Library managers can use this iDSS tool to ensure that different perspectives are taken into account in a decision-making process. In addition, the iDSS provides the data-based justifications for managerial and economic decisions library managers must make.

The DW implementation in CDRJBV not only demonstrates how to integrate strategic information but also allows improving library performance through more mature transactional processes that take place daily. For instance, during the data analysis phase, a number of inconsistencies were detected in the bibliographical database, mainly due to typographical errors made by catalogers. Consequently, the DW implementation strengthens the importance of using standards, policies and procedure manuals for data entry in order to reduce time consumption during the data cleaning process. Some of the challenges encountered when implementing the DW are the lack of integration between library and university systems and the high cost and time of loading initial data into the data warehouse.

Finally, it is important to consider that the benefit of the DW increases when data analysis techniques are considered in the iDSS architecture. Strategic data stored in the data warehouse can be used for different purposes: 1) data visualization and reporting, allowing library manager to publish library indicators in a simple and quick manner by using online reporting tools; 2) sophisticated data analysis through the use of data mining tools; and 3) input for optimization models. Data mining techniques analyze large information databases and discover implicit, but potentially useful information. Data mining has the capability to uncover hidden relationships and to reveal unknown patterns and trends by digging into large amounts of data. The appropriate use of data mining techniques and optimization models in libraries is the next step after the DW implementation, and is analyzed in RQ4.

10.2.4 Data Analysis and Presentation

RQ4: What tools and strategies can be used to visualize and analyze strategic information to support libraries in decision-making?

The application of data mining techniques in libraries is an emerging trend that has captured the attention of practitioners and academics in order to understand patterns of behavior of library users and staff, and patterns of information on resource usage throughout the library. In order to understand the use of data mining techniques in libraries, a comprehensive literature review was

carried out in Chapter 7. Forty-one papers were identified, analyzed and classified along the four quadrants of the holistic evaluation matrix and the main data mining functions, which are, clustering, association, classification, and regression. Relevant insights and recommendations on the state-of-the-art of data mining tools in libraries are highlighted. For instance: the importance in implementing more data mining functions to support holistic-based decision-making process; the need to conduct further research in the area of quality control and in the analysis of service performance in both digital and physical environments; the possibility to implement unsupervised algorithms such as association and clustering models, techniques highly recommended for non-experts in data mining.

Based on the findings of this comprehensive literature review and the iDSS architecture implemented in the CDRJBV, an experimental use of data mining techniques for library decision-making was addressed in Chapter 8. In particular, algorithms of regression, clustering and classification have been applied to the case study in order to predict future investments in library development, to find clusters of users that share common interests and similar profiles but belong to different faculties, and to establish possible correlations between student library collection usage and their final results taken on a pass/fail basis.

The first exercise addresses a strategic problem as is the anticipation of future investments in library development. To do so, three regression techniques are used. During data cleansing, several outliers were detected and replaced. These outliers are the result of political changes and law enforcements. Findings indicate that the three models behaved very similarly and were able to forecast the time series direction of change, e.g. increase or decrease trends; as well as, the magnitude of the predicted values. Although, the study shows promising results to forecast library investment, there is still room for improvements on the predictions, since time series depend on different exogenous factors such as sporadic donations, political changes and law enforcements that for this experimental exercise were not included in the predictive models. For instance, budget for the library in this university is allocated by faculties; CDRJBV does not manage its own funds. Thus, each faculty decides what to subscribe and what to unsubscribe to based on their own finances, priorities, and political decisions. Therefore, an enhancement to improve model accuracy would be to incorporate variables that reflect expected political or policy changes. It was also observed that there was not a trend in the yearly expenses values but an intra-year pattern. There were months with larger expenses such as March, October, November and December. The rationale of these results is that large expenses occur at the beginning of each academic semester (March and October) and close to the end of the fiscal year (November, December). These preliminary prediction results can be used as a yearly indicator on how much expenses could be expected. Furthermore, for CDRJBV, these “encouraging” results can be used to justify the need for increasing autonomy, particularly the need for controlling their own budget.

The second exercise finds clusters of users that belong to different faculties but share common interests and similar profiles such as academic performance, socio-economic aspects, and knowledge areas of the borrowed books. These results can be used as input for the creation of new library policies, as well as keep users with similar interests informed of new resources on specific topics. In this clustering exercise, five algorithms were first tested; then two techniques were discarded after initial proving. Finally, based on the analysis of the visual representation of the cluster distribution, a clustering output was selected as the most promissory. An intra-cluster analysis was performed in the most interesting cluster that consisted of users registered in different faculties, and where the number of users is representative enough. Findings indicate that this cluster of users mainly coming from the Faculties of Medicine, Engineering, and Humanities are heavily interested in technical subjects. A cluster like this provides the input for a more in-depth analysis in order to ensemble, for instance, a multidisciplinary team to perform research activities.

Eventually, the third exercise predicts student academic performance based in factors such as library usage, academic and socio-economic profiles, by means of a classification task. Preliminary findings indicate that no notable associations were found among socioeconomic background, library usage, and academic performance; although, family income factors may affect student academic performance, since it was found that library users with higher income have more chances

to pass. For the institution, these type of findings can support the development of library services to target specific student groups on the basis that higher library usage may lead to improved academic performance. In addition, this experimental study describes a research design that is replicable in other libraries and contributes to library usage and learning analytics literature. In addition, these preliminary findings provide basis for further investigation on this topic and demonstrate how institutional data can be combined to examine library usage and academic performance at a single institution. Undergraduate students and library usage data were analyzed to identify results that suggested associations or relations between library usage and academic performance.

Despite its experimental character, the study shows that the use of data mining techniques under a holistic approach can provide the data-based justification for their identified library needs, or the appropriateness of particular decisions. The most important implication of this exercise is that having as a baseline the holistic framework and the iDSS architecture, the study demonstrates how multiple measures and data mining techniques are able to be applied to the integrated framework in order to understand the entire library system. There are, however, some considerations that have to be taken into account such as the quality and effectiveness of the results which depend, to a great extent, on the availability of relevant information and the quality of data. The more data are provided, the more accurate results will be obtained.

Finally, early experiences developing an optimal resource allocation model for distributing resources among different processes of a library system utilizing a holistic framework as data input is addressed in Chapter 9. This preliminary approach presents several challenges. The first main challenge to be faced is how to quantify usage. To do so, a combination of methodologies coming from the third quadrant of the holistic matrix is recommended, that is, citation analysis, vendor-supplied statistics, and citation databases. To this end, some assumptions are made: a percentage of increase over the usage values is utilized as the basis to calculate the future usage; and the number of publications per journal is considered more important than the number of citations and vendor-supplied statistics. The second challenge is to combine the data coming from citations, publications, vendor-supplied statistics, and journal rankings as a measure of usage. Ideally, a general strategy is to assign weights to each number and use the combined data in the model; thus, each subject area can have its own weight. However, assigning weights to each subject area is extremely difficult due to the large number of these areas. To overcome these two first challenges, journal titles are first grouped based on subjects' areas or categories, and then these categories are divided into four subcategories based on the quartile rankings of the journal citation reports (JCR) database. These subcategories are used in the objective function to make decisions on the number of journals to be purchased per subcategory. Finally, an additional challenge that arises when categorizing journals when using the JCR classification, is that journals can be assigned to more than one category; therefore, a strategy to avoid double-counting is required. In addition, this study only discusses how a library should manage the budget for journals. The expenses incurred on books, salaries, maintenance, and other indirect costs were not considered in this initial model. Although this is just an initial approach toward a complete solution, we may conclude that this optimization technique seems feasible, with potential benefits for library managers. In addition, this stage of the system allows researchers to identify opportunities for future studies and applications. The implementation of the complete solution based on the holistic approach is undoubtedly a must for future research.

10.3 Optimal Resource allocation and Budgeting in Libraries. The ORBIL approach

In order to ensure replicability of results and techniques, this section summarizes the ORBIL approach by describing the activities, deliverables, best practices and indicators that can be used to replicate the proposed approach in other library settings by other researchers and eventually practitioners. The ORBIL approach, as depicted in Figure 10.3, proposes to divide the framework into three big blocks: 1) data source (Table 10.1); 2) data storage (Table 10.4); and 3) data analysis and presentation (Table 10.5).

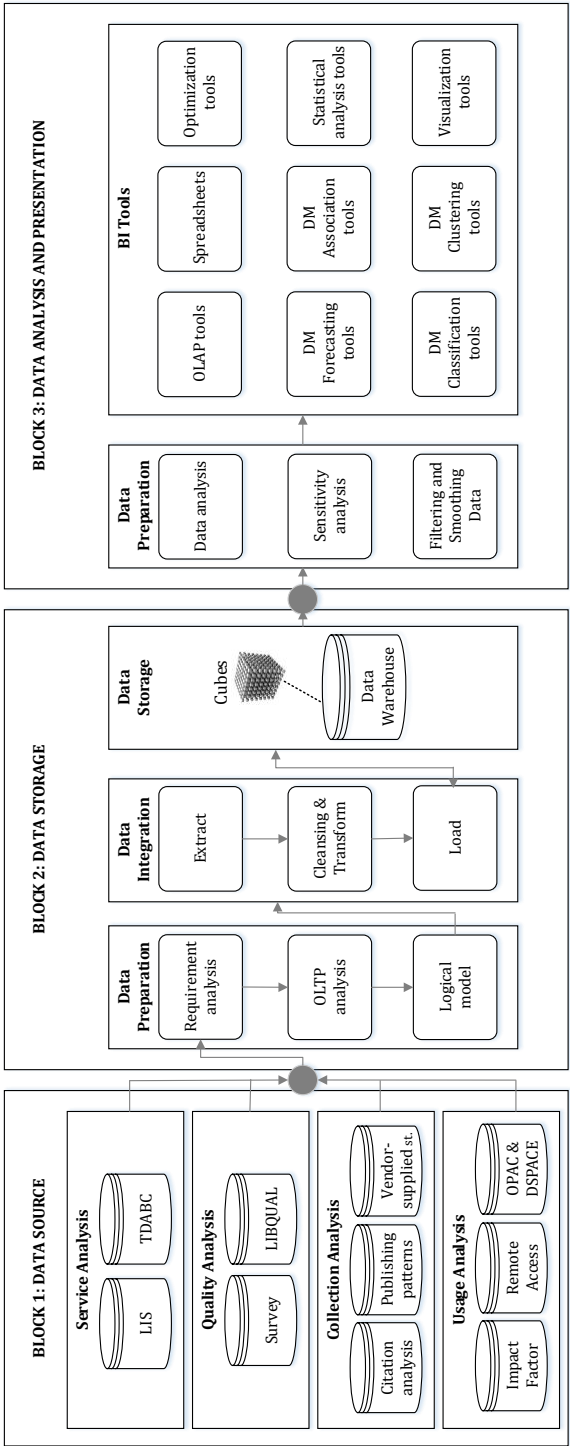


Figure 10.3: ORBIL Framework

Table 10.1: Description of the first block, including activities, deliverables, best practices, indicators and future work

BLOCK N°	1
TITLE:	DATA SOURCE
<p>Objectives and approach:</p> <p>The aim of this work package is to ensure data collection performed in a structured manner, covering the key aspects of a library, and at the same time, facilitating the understanding and replication of the data collection process</p> <p>Recognizing the need to evaluate libraries in a holistic and structured manner, this study utilizes a two-dimensional evaluation matrix. This holistic framework evaluates the library system and collection from an internal and an external perspective. The approach for implementing this matrix starts by identifying the services or activities involved in libraries and by calculating the costs of different resources (staff, equipment, facilities, collection, etc.). In order to do so, qualitative mechanisms for assessing library effectiveness are included -for example, observation, interviews, surveys, expert opinions, process analysis, organizational structure analysis, standards, and peer comparison. Quantitative techniques are also required to evaluate efficiency, usefulness, and manipulation of the system. Citation analysis, log analysis, statistics gathering, and stopwatch techniques are useful methods to be included.</p>	
ACTIVITY N°	1
TITLE:	SERVICE ANALYSIS
<p>The processes and services carried out within the library system are the main aspects studied. From an economic point of view, the ORBIL approach analyzes the costs incurred and the resources consumed by the library processes through the use of Time-Driven Activity-Based Costing (TDABC).</p> <p>Subactivities:</p> <ul style="list-style-type: none"> • Identify services, processes and resource groups • Estimate the total cost of each resource group • Estimate the practical time capacity of each resource group • Calculate the unit cost of each resource group • Determine the estimated time for each activity • Calculate the unit cost of each resource group • Analyze results and develop strategies for improvement on the basis of the analysis. 	
<p>Deliverables:</p> <ul style="list-style-type: none"> • Summary of the services, processes, activities and resource groups involved in the library system • Report of unit costs of services, processes, activities and resource groups • Process flows diagrams including activities, costs and time duration • Time equations per library process 	
<p>Best practices:</p> <ul style="list-style-type: none"> • Creating graphical representations of activity flows allows validating the collected information straightforwardly because librarians can easily understand the sequences and their responsibility in each process. To this end, two types of tools are suggested: 	

<ul style="list-style-type: none"> ○ Combining MS Office suite programs, such as Visio and Excel ○ Implementing a dedicated software tool, such as TD-ABC-D in order to keep the model updated and consequently to facilitate long-term maintenance • Data collection on the duration of activities needs to be gathered by direct observation and a stopwatch. The most accurate data are collected when librarians physically perform the tasks, although this type of data collection takes a significant amount of time. • Involvement and commitment of the library staff are critical to the data collection by increasing the acceptance of the model. Some staff members may feel uncomfortable being observed while working. This discomfort can cause some resistance and consequently can delay data collection. Proper communication as well as the involvement and commitment of the managers and staff can increase the level of acceptance. In addition, library managers and TDABC team should explain the purpose of the measurement, importance of the model, activities to perform, and implications of the results. They should clearly state that the activities and profiles are measured, not the names of individuals. Therefore, motivation and an explanation of the measurement purpose are fundamental for achieving the desired commitment with staff. 	
Indicators: <ul style="list-style-type: none"> • Number of library processes documented/total number of library processes 	
Future work: <ul style="list-style-type: none"> • Implementation of TDABC in activities that follow unstructured and non-systematic sequences such as user reference processes. 	
ACTIVITY N°	2
TITLE:	QUALITY ANALYSIS
<p>In this activity, the users' perception about the quality of the offered services is evaluated. The ORBIL approach recommends the use of LibQUAL+®, since this is the most popular and widely used survey instrument. Other assessment methods include: statistics gathering, suggestion boxes, web usability testing, user interface usability, and satisfaction surveys.</p> <p>Subactivities:</p> <ul style="list-style-type: none"> • Set the survey according to the library requirements and expectative <ul style="list-style-type: none"> ○ Language ○ Population ○ Local questions • Perform the LibQUAL+® survey with students and staff • Analyze the survey results 	
<p>Deliverables:</p> <ul style="list-style-type: none"> • Overview of the LibQUAL+® results: minimum, desired, perceived, adequacy gap, superiority gap, and number of subjects <ul style="list-style-type: none"> ○ Per group of users ○ Per library branch ○ Per discipline ○ Per gender • Report of the main strengths and weaknesses discerned from the analysis • Statistics and benchmark results with other library branches and peer institutions 	

Best practices:

- The deployment period requires special attention in order to avoid possible mistakes and delays; for example, setting the language and population.
 - Although LibQUAL+® provides the standard questionnaire in different languages, in some cases, it is necessary to make some specific changes that have to be coordinated together with the Association of Research Libraries, for instance, when chosen to apply the survey in two languages.
 - Gathering the population data for each user group is not an easy task. In the case of students, for instance, it can be important to distinguish the year that the student is being trained and not the year that the student is registered.
 - Determining the number of PhDs by discipline requires special consideration because of the multidisciplinary groups.
- A personalized tool is required in order to analyze each library branch and discipline, since these types of results are not provided by standard reports. The LibQUAL+® survey produces standard reports in which the measurement is carried out in its overall performance and user groups (e.g. students, PhD's, faculties). This standard report also provides no direct insight into the library performance compared to other libraries branches where the LibQUAL+® survey is performed.
- Involvement of the University is crucial to obtain good participation rates and results. Although LibQUAL+® Lite improves response rates and reduces respondent burden, in some cases, there is still the perception that the survey is very long. Other strategies to stimulate the users' response are by offering incentives, such as electronic devices and movie tickets, as well as by sending a reminder email two weeks before the close of the survey. In addition, an input field for free text gives users the opportunity to submit comments regarding their concerns and to express suggestions for future improvements.
- LibQUAL+® is a standardized tool; however, a benchmarking study should be interpreted with extreme caution because the survey compares what the users think about libraries and not about what the library really is.

Indicators:

- User response rate

Future work:

- Implementation of personalized surveys and quality assessment methods required based on the holistic framework approach.

ACTIVITY N°

3

TITLE:

COLLECTION ANALYSIS

This activity allows quantifying the impact of the library collection on its users, providing library managers with better basis for decision making when acquiring new bibliographic materials. In order to accomplish this, the ORBIL approach combines citation analysis, citation database and vendor-supplied statistics. The aim of this activity is to provide a deep insight of the local use of the library collection.

Subactivities:

- Citation analysis of PhD theses.
- Publishing pattern analysis of students, academic staff and/or researchers.
- Analysis of vendor-supplied statistics of all journals downloaded during a specific period and subscription costs.

Deliverables: <ul style="list-style-type: none"> • Database of journal citations • Database of publishing patterns • Database of journals downloaded during a specific period • Overview of the most popular journals utilized by library users <ul style="list-style-type: none"> ○ Per faculty ○ Per department ○ Per individual user • Overview of subscription costs 	
Best practices: <ul style="list-style-type: none"> • An automated tool is highly recommended since an average of 2.5 hours is necessary to manually analyze a thesis in order to both collect the information and incorporate them in the different databases. In addition, this dedicated software will help to collect the large amount of information, as well as to speed up the evaluation of the results. • Mapping abbreviations can help to reduce time of analysis. In literature, there is no defined standard for journals' abbreviations and acronyms, thus collecting journals' information is not always straight forward. For instance, the ISO Abbreviation for the Journal of the American Chemical Society is J. Am. Chem. Soc.; the JCR Abbreviation is J AM CHEM SOC, while its acronym is JACS. Proc. IEEE is the ISO abbreviation for The Proceedings of the IEEE and its JCR abbreviation is P IEEE. Therefore, a certain expertise is necessary to differentiate the different abbreviations and acronyms of journals that students cited as reference. 	
Indicators: <ul style="list-style-type: none"> • Number of theses analyzed vs. total number of theses 	
Future work: <ul style="list-style-type: none"> • Implementation of a module for automated citation analysis and publishing patterns • Implementation of a reference table with abbreviations 	
ACTIVITY N°	4
TITLE:	USAGE ANALYSIS
<p>This activity analyzes the use patterns followed to manipulate the library system. For instance, in digital library services, it is possible to track everything users search and retrieve from the library system. To analyze this users' behavior, the ORBIL approach relies on usage statistics and log analysis methods.</p> <p>Subactivities:</p> <ul style="list-style-type: none"> • Collect and analyze log files, filtering the older records out from consideration • Perform a log analysis • Collect usage statistics from the library information system. • Analyze the usage statistics, aggregating the results per month, semester, and/or year. 	
Deliverables: <ul style="list-style-type: none"> • Database of log results • Overview of usage statistics <ul style="list-style-type: none"> ○ Physical library ○ Digital library 	

Best practices:

- Encode the user identification by replacing the user ID with a code in order to avoid privacy infringement. Another option is to create a demographic surrogate to replace personal information about the user through a set of demographic values (e.g. age, sex, education).
- Carefully define the range of IP addresses to be monitored, since in most cases, the academic library is part of a University system

Indicators:

- Number of usage indicators available
- Number of register collected

Future work:

- Implementation of a dedicated web log analysis

Some additional considerations to be taken into account when collecting data from heterogeneous data sources include the following:

- Lack of well-defined standards for some specific analysis, such as the abbreviation of journal names, access to electronic collection, and e-lending.
- Need for a common understanding of what sources and data must be considered.
- Need for integrating multiple data sources from the library, university, consortiums, and suppliers.
- Differences of requirements between traditional and digital collections (for example, digital libraries require licenses for a certain time period, links to remote resources, or prepaid pay-per-view).
- Large volume of data generated by all different sources, for instance, web logs.

Examples of the implementation of the first block for the book analysis is presented in Table 10.2, and for Circulation and ILL services in Table 10.3.

Table 10.2: Example of the first block of the ORBIL approach applied to books

Services	Processes	Process Analysis	Quality Analysis	Collection Analysis	Usage Analysis
WBIB (Open shelve collection)	WBIB selection	WBIB costs	IC-3 – The printed library materials I need for my work	# books cited # books published	# WBIB books # WBIB books acquired # WBIB books cataloged # WBIB books consulted # linear meters of WBIB shelving # total staff for WBIB cataloging (FTE) # total staff for WBIB acquisition (FTE) # students assistants for WBIB # WBIB items repaired
	WBIB acquisition	Copy cataloging process costs			
	Copy cataloging	Acquisition process time			
	Holding cataloging	Copy cataloging process time			
e-books	e-book selection	e-book selection costs	IC-4 – The electronic information resources I need	# e-books cited # e-books published # e-books downloaded	# e-books # e-books added # e-books consulted
	e-book acquisition	e-book acquisition costs			
		e-book selection time			
		e-book acquisition time			
WMAG (Stack collection)	Copy cataloging	Copy cataloging costs	IC-3 – The printed library materials I need for my work	# books cited	# WMAG books # WMAG books added # WMAG books cataloged # WMAG books consulted # WMAG book records deleted # linear meters of WMAG shelving # total staff for WMAG cataloging (FTE) # students assistants for WMAG # WMAG items repaired/preserved
	Holding cataloging	Holding cataloging costs			
	Original cataloging	Original cataloging costs			
	Deleting records	Deleting records costs			
WAIT collection = a book staying in a particular department	Copy cataloging	Copy cataloging costs	IC-3 – The printed library materials I need for my work	# books cited	# WAIT books # WAIT books added # WAIT cataloged # WAIT books consulted # total staff for WAIT cataloging (FTE)
	Holding cataloging	Holding cataloging costs			
		Copy cataloging time			
		Holding cataloging time			
WDEP collection (External Stack collection)	Copy cataloging	Copy cataloging costs	IC-3 – The printed library materials I need for my work	# books cited	# WDEP books # WDEP books added # WDEPbooks cataloged # linear meters of WDEP shelving ...
	Holding cataloging	Holding cataloging costs			
	Original cataloging	Original cataloging costs			
		Copy cataloging time			
		Holding cataloging time			
		Original cataloging time			

Table 10.3: Example of the first block of the ORBIL approach applied to Circulation and ILL services

Services	Processes	Process Analysis	Quality Analysis	Collection Analysis	Usage Analysis
Lending	WBIB Lending	WBIB Lending costs WBIB Lending time	IC-5 – Modern equipment that lets me easily access needed information IC-7 – Making information easily accessible for independent use	# WBIB cited # WBIB published	# WBIB Lendings # Lending machines
Returning	WBIB Returning	WBIB Returning costs WBIB Returning time	IC-5, IC-7		#WBIB Returns # returning machines # students assistants for Returning
Hold request	WMAG Lending	WMAG Lending costs WMAG Lending time	AS – Affect of Service IC-5 LOCAL-3 – Efficient interlibrary loan / document delivery	# WMAG cited	# WMAG Lendings # lending machines # students assistants for WMAG request
Interlibrary Loan	ILL Outgoing request – book	ILL Outgoing request costs ILL Outgoing request time	AS, LOCAL-3 LOCAL-7 – Easily and quickly obtain materials from other libraries.	# books published	# ILL - Outgoing request - book # computers available for ILL # total staff for ILL (FTE) # students assistants for ILL
Interlibrary Loan	ILL Outgoing request – journal	ILL Outgoing request costs ILL Outgoing request time	AS, LOCAL-3, LOCAL-7	# journal articles published	# of ILL - Outgoing request - journal # of computers available for ILL # of total staff for ILL (FTE) # of students assistants for ILL
Interlibrary Loan	ILL Incoming digital	ILL Incoming digital costs ILL Incoming digital time	AS, IC-7, LOCAL-3, LOCAL-7	# journal articles cited	# of ILL - Incoming Request - journal online # of computers available for ILL # of total staff for ILL (FTE) # of students assistants for ILL
Interlibrary Loan	ILL Incoming printed	ILL Incoming printed costs ILL Incoming printed time	AS, IC-7, LOCAL-3, LOCAL-7 IC-8 - Print and/or electronic journal collections I require for my work	# books cited	# of ILL - Incoming Request - book printed # of computers available for ILL # of total staff for ILL (FTE) # of students assistants for ILL

Table 10.4: Description of the second block, including activities, deliverables, best practices, indicators and future work

BLOCK N°		2
TITLE:		DATA STORAGE
<p>Objectives and approach:</p> <p>With a complete framework for data collection, the data collected in Block 1, coming from multiple sources, and therefore with different formats, need to be integrated and stored in an adequate structure for decision support. Subsequently, such solution should allow data manipulation, analysis and visualization. Unfortunately, this integration presents a big challenge, since these different data sources normally use dissimilar formats and access methods. The objective of this second block is to implement a DW to integrate, filter and process all the information extracted from many different systems based on a holistic approach.</p> <p>Building a DW involves extracting data from different data sources, in which many problems of inconsistency need to be dealt with. It also involves a process of data extraction, cleansing and storage through ETL (Extract, Transform, Load) processes. This process is complex and time-consuming, because it needs to combine all the different data sources and converts them into a uniform format, excluding possible inconsistencies, redundancies, and incompatibilities.</p> <p>The architecture chosen for the DW implementation is Hefesto (Chapter 6). This methodology allows tackling the design of the DW from different detail levels, and reducing risks of failure and dissatisfaction by involving end-users early in the design process. The Hefesto methodology starts by identifying user information needs to define all queries of interest. Next, a data source analysis is performed in order to determine how the indicators are built, to define correspondences and granularity, and to build the extended conceptual model. A logical model that represents the structure of the DW is then defined to set the type of implementation schema, the dimension and fact tables. Eventually, a diverse set of tools, such as cleansing techniques, data quality control, and ETL processes are utilized in order to integrate the data of different data sources, policies and strategies.</p>		
Activities:		
A1	<p>Requirement analysis:</p> <p>Based on the holistic evaluation framework for data collection implemented in Block 1, a set of queries of interest to be issued against the DW is defined in this activity. This list of requirements is collected through questions on library needs. The requirements should be documented, actionable, measurable, testable, traceable, related to identified library needs or opportunities, and defined to a level of detail sufficient for system design.</p> <p>In order to define the requirements, rounds of interviews with the library manager and stakeholders are required. Other techniques that can be used in this activity include the development of scenarios, the identification of use cases, the use of direct observation, and creating requirements lists. Where necessary, the requirement analyst can employ a combination of these methods to establish the exact needs of library managers.</p> <p>Subactivities:</p> <ul style="list-style-type: none"> Identify questions posed. The main objective of this task is to obtain and identify the key information needs. Identify indicators and perspectives. Indicators are normally numerical measures, while perspectives are related to objects through which, it is required to examine the indicators. Build the conceptual model 	

<p>A2</p>	<p>OLTP analysis:</p> <p>The following activity in the second block identifies the different data sources of the library based on the requirement analysis of the holistic evaluation approach. The holistic approach incorporates several key elements including process analysis, quality estimation, information relevance, and usage interaction; thus, data have to be collected from internal and external sources. Internal data sources refer to the databases that are managed at the library level. On the contrary, external sources are not managed by the internal processes of the library.</p> <p>In this activity, the OLTP sources are analyzed in order to determine how the indicators are calculated, and to establish the respective correspondences between the conceptual model created in the previous step and data sources. Then, the fields that should be included in each perspective are defined. Finally, the conceptual model is expanded with information obtained in this step.</p> <p>Subactivities:</p> <ul style="list-style-type: none"> • Establish indicators. This task requires to explain how the indicators are calculated. • Establish correspondences between the conceptual model and data sources. • Select the fields that contain each perspective. • Expand the conceptual model, including indicators, fields and correspondences.
<p>A3</p>	<p>Logical Model:</p> <p>In this activity, a data mapping from the OLTP sources to the logical model is performed based on the conceptual data model. Moreover the type of schema is defined, such as star, snowflake or fact constellation. By selecting the type of schema, the dimensions and fact tables are built, and then joint to create multidimensional models. In a star schema, for example, facts are the core data elements being analyzed, and the dimensions are the attributes about facts.</p> <p>Subactivities:</p> <ul style="list-style-type: none"> • Define the type of scheme to be used • Design the dimension tables that will be part of the DW • Define fact tables • Create the respective unions between dimension and fact tables.
<p>A4</p>	<p>Data integration:</p> <p>After building the logical model in Activity 3, the relevant data generated by multiple sources are extracted and integrated by means of cleansing techniques, data quality control, and ETL processes. This allows having a clean and homogeneous version of the library data. Because this process is the most tedious and time-consuming part, it is recommended starting with a narrowly specific query and working through the entire process, and then, iteratively continuing developing the DW.</p> <p>Subactivities:</p> <ul style="list-style-type: none"> • Extracting data from transactional and documental databases. • Cleansing and transforming data. Some of the common tasks of this subactivity are: filtering data, converting codes, calculating derived values, transforming between different data formats, and automatic generation of sequence numbers. • Loading data. The transformed data are loaded into the DW.

Deliverables:

- List of requirements
- List of indicators and perspectives
- Summary of library data sources
- Conceptual data models
- Multidimensional models

Best practices:

- An adequate selection of methodology and technological tools for constructing the DW is necessary to ensure the data warehousing success. For instance, the use of a DW architecture and modeling at early deployment time helps library managers and stakeholders to realize the potential of implementing a DW solution in order to make tactical decisions.
- The use of standards, policies and procedure manuals for data entry reduces time consumption during data cleaning process.
- The integration between library and university systems must be considered not only for the DW, but also for operational systems.
- Initial data loads must be planned carefully as certain tables may need to be loaded first to help verify loads of other tables. It is usually better to perform the initial loads incrementally and by establishing a detailed strategy in order to reduce the high cost and time of loading initial data.
- The more historical data are available; the more helpful reports will be obtained.
- The benefit of the DW increases when data analysis and data mining techniques are considered in the full approach.

Indicators:

- Number of multidimensional cubes created
- Number of data sources integrated in the DW vs. total number of data sources available in the library

Future work:

- Implementation of a personalized tool for DW design.
- Incorporation of additional data sources, such as the syllabus management system, citation analysis, and Web portal statistics
- Use of semantic technologies to integrate potential data sources.

Table 10.5: Description of the third block, including activities, deliverables, best practices, indicators and future work

BLOCK N°		3
TITLE:		DATA ANALYSIS AND PRESENTATION
<p>Objectives and approach:</p> <p>Strategic data stored in the DW can be used for different purposes: 1) data visualization and reporting, allowing library managers to publish library indicators in a simple and quick manner by using online reporting tools; 2) sophisticated data analysis through the use of data mining tools; and 3) input for optimization models. The aim of this block is to implement tools and techniques that can be used to visualize and analyze strategic information to support libraries in decision making.</p> <p>The approach for implementing this block starts by integrating and cleansing data in order to remove outliers, duplicates and inconsistencies. These cleansed data are then transformed into appropriated formats that can be understood by data mining tools and optimization techniques, and filtration and aggregation techniques are applied to the data in order to extract summarized data. In fact, interesting knowledge is extracted from the transformed data. This information is analyzed in order to identify the truly interesting patterns. Eventually, knowledge is visualized to library managers.</p>		
Activittes:		
A1	<p>Data mining analysis:</p> <p>Data mining techniques analyze large information databases and discover implicit, but potentially useful information. Data mining, as any knowledge extraction method, follows a systematic procedure to allow an appropriate knowledge discovery. The data mining process starts by determining areas of focus and collecting data; then, these data collected are cleansed, and anonymized. To discover meaningful patterns in the collected data, the data mining process includes the selection of appropriate analysis tools and data mining techniques. Interesting patterns are analyzed and visualized through reports. The mining process needs to be iterated until the resulted information is verified and proved by key users such as librarians and library managers.</p> <p>Subactivities:</p> <ul style="list-style-type: none"> • Prepare data for mining <ul style="list-style-type: none"> ○ Detect and remove noise or outliers ○ Smooth the signal ○ Apply aggregation or filtering techniques • Analysis of the most appropriated technique to be applied in order to match the analysis goals • Build the data mining model • Evaluate the model's accuracy and performance • Visualize results 	
A2	<p>Optimization model:</p> <p>Similar to the previous activity, optimization techniques start the process preparing data to be utilized in the optimization model. Then, traditional optimization techniques define a</p>	

	<p>potential solution to a problem, and then improve them iteratively.</p> <p>Subactivities:</p> <ul style="list-style-type: none"> • Define scenario • Analyze data requirements: indices, decision variables, parameters and possible constraints • Prepare data for data input • Select the optimization technique • Formulate the optimization model. Define the objective function and constraints • Deploy the model • Evaluate the model. For instance, perform a sensitivity analysis of the results to evaluate the robustness of the solution • Visualize results
	<p>Deliverables:</p> <ul style="list-style-type: none"> • Report of the data mining results. For instance, predictive results, clustering and classification groups, possible associations and relations. • Optimized scenarios, description of the models and results
	<p>Best practices:</p> <ul style="list-style-type: none"> • Graphical representation of results provides many visual insights to help managers to validate the data mining process as well as to interpret the results. • Successful data mining implementations in libraries require the involvement of expertise in library management, library data and data mining, techniques. Library management knowledge is required in order to establish the requirements and interpret results. Library data expertise is necessary to facilitate data collection and preparation. Eventually, data mining expertise is required to interpret the library needs, to select the appropriate analyses and data mining techniques to be utilized, and support the interpretation of results. • The use of established standards for data collection and analysis increases acceptance of the process among library managers, staff and institutional authorities. It also increases compatibility with local library systems and other systems' institutions. • Decision making cannot be fully captured by standard models or approaches, a certain degree of autonomy and subjectivity need to be incorporated in the model.
	<p>Indicators:</p> <ul style="list-style-type: none"> • Percentage of reduction in costs, time and resources • Percentage of increase in service quality • Percentage of services improved • Number of data mining functions deployed • Number of data sources/quadrants covered by the data mining functions • Number of data sources/quadrants utilized in the optimization model
	<p>Future work:</p> <ul style="list-style-type: none"> • Deploy data mining applications to a production environment • Deploy optimization modules to an experimental and production environment

10.4 Overall conclusion

Budget allocation is a core problem faced by all academic libraries independent of their size and funding mechanism. Although resource allocation is a complex process, it is ever more necessary especially in environments of constant change and budget adjustments. The main purpose of this study is to develop an integrated model that can support libraries in making optimal budgeting and resource allocation decisions among their services and collection through a holistic analysis. To this end, a combination of several methodologies and structured approaches is conducted. Firstly, a holistic structure and the required toolset to holistically assess academic libraries are proposed to collect and organize the data from an economic point of view. Secondly, a data warehousing approach is recommended and implemented to integrate, process, and store the holistic-based collected data. Ultimately, several techniques are explored and tested to visualize and analyze the stored data that can help libraries in their decision-making, such as reporting and data mining tools, and optimization models. By proposing this holistic approach, this research study hopes to contribute knowledge by providing an integrated solution to assist library managers to make economical decisions based on an “as realistic as possible” perspective of the library situation.

*“Allocating ... budget must be an act of balancing limited resources against
seemingly limitless needs.”*

(Wise 1996)

Appendices

Appendix A: **Towards a holistic Analysis**

Tool to support decision-making in

libraries

Siguenza Guzman, L., Holans, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., Cattrysse, D. (2013). Towards a holistic analysis tool to support decision-making in libraries. Proceedings of IATUL Conferences. 34th Annual IATUL Conference. Cape Town, South Africa, 14-18 April 2013 (art.nr. 29) (pp. 1-9) Purdue e-Pubs.

Abstract

Academic libraries have recently been subjected to continuous budget reductions, mainly due to the increasing costs of information and the global economic crisis. As the primary purpose of an academic library is to provide well-balanced collections and a wide range of services to support education and research, an efficient use and allocation of limited resources is vital. However, allocating resources such as money, staff, time, and infrastructure between the library collection and services represents a challenge due to the multitude of data sources required to consult during a decision-making process.

Academic libraries are accustomed to keeping voluminous statistics on their collection and services; however these data are not fully used for decision-making processes due to the lack of an efficient structure for grouping this information. The authors in a previous study state that prior to decision making, data must be collected based on a holistic approach that incorporates all of the key elements that may influence a decision. It is in this sense that to holistically assess libraries, an approach combining a theoretical framework with several measurement tools is proposed in that study. Therefore, the aim of this paper is to document early experiences and lessons learned in implementing the holistic approach in an academic library in Belgium. To do so, the academic library is evaluated in two dimensions. The first dimension analyzes the library system and its collection, whereas the second dimension analyzes the perspective of both the user and the internal stakeholders. During the initial implementation stages, the proposed approach proved to be valuable to ensure a complete view of the library collection and services. There are, however, important considerations to be borne in mind such as the time required to implement the complete approach, as well as the need of a system to integrate the collected information.

A.1 Introduction

Amid limited funding resources, academic libraries are striving to efficiently satisfy the growing demands for new and flexible services. David J. Ernst and Peter Segall (1995) state that institutions in these difficult circumstances are called to develop a strategic and well-coordinated budget plan by means of a “holistic approach”. This holistic approach requires interconnecting all necessary components in a way that responds to both, shrinking resources and dynamic library services.

Academic libraries are accustomed to collecting statistics about their collection and services. However, these data are not fully utilized for decision-making processes due to the lack of an efficient methodology for grouping and analyzing this information. The authors in a previous study¹⁸ (2013) proposed an approach which combines a theoretical framework with several measurement tools to holistically assess libraries prior to decision making. The goal of this paper is to highlight the key benefits, challenges and lessons learned in implementing the proposed holistic approach in an academic library in Belgium.

A.2 Theoretical background

Holism is a concept which emphasizes the importance of the whole and the interdependence of its parts (Editors of the American Heritage Dictionaries, 2011). If this concept is applied to libraries, it can be interpreted as an analysis that emphasizes the importance of the entire library and the interdependence of its processes, collection and services. In this respect, Lorena Siguenza-Guzman, Alexandra Van den Abbeele, Joos Vandewalle, Henri Verhaaren, and Dirk Cattrysse (2013) propose a holistic approach to be used prior to developing a budget plan. The approach combines the theoretical framework proposed by Scott Nicholson (2004), with several evaluation tools. This framework shown in Figure A.1 uses a two-dimensional evaluation matrix, in which columns represent the topic (library system and collection), and rows represent the perspective (library staff and users).

		Topic	
		Library System	Use
Perspective	Internal (Library System)	1. What does the library system consist of?	4. How is the library system manipulated?
	External (Users)	2. How effective is the library system?	3. How useful is the library system?

Figure A.1: Conceptual matrix for holistic measurement (Nicholson, 2004)

The following paragraphs briefly describe the main features of each quadrant based on the holistic approach:

- *First quadrant: internal perspective of the library system.* The processes and services carried out within the library system are the main aspects studied. From an economic point of view as required in this study, Siguenza-Guzman et al. propose to analyze the costs incurred and

¹⁸ Updated reference: Siguenza Guzman, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., Cattrysse, D. (2015). A holistic approach to supporting academic libraries in resource allocation processes. *Library Quarterly*, 85 (3).

the resources consumed by the library processes through the use of Time-Driven Activity-Based Costing (TDABC).

- *Second quadrant: external perspective of the library system.* This quadrant evaluates the users' perception about the quality of the offered services. To do so, Siguenza-Guzman et al. recommend the use of at least one of the top five assessment methods reported by Stephanie Wright and Lynda S. White (2007). These methods are statistics gathering, suggestion boxes, Web usability testing, user interface usability, and satisfaction surveys.
- *Third quadrant: external perspective of use.* This quadrant allows quantifying the impact of the library collection on its users, providing library managers with better basis for decision making when acquiring new bibliographic materials. In order to accomplish this, Siguenza-Guzman et al. propose to combine citation analysis, citation database and vendor-supplied statistics.
- *Fourth quadrant: internal perspective of use.* The fourth quadrant analyzes the use patterns followed to manipulate the system. For instance, in digital library services, it is possible to track everything users search and retrieve from the library system. To analyze this users' behavior, Siguenza-Guzman et al. propose to incorporate log analysis methods such as transaction log analysis and deep log analysis.

A.3 Holistic analysis tool to support decision-making in libraries: Case study

A.3.1 The case study

A case study was conducted at the Arenberg Campus Library (CBA - Campusbibliotheek Arenberg) of the KU Leuven in Belgium. The CBA staff, approximately 19 full-time equivalent employees (FTE) provide service to about 10,000 potential customers. To improve cost efficiency and effectiveness, the CBA has been forced to find new strategies to deliver its services, such as the use of new technologies, improving access to e-journals and databases, automation of repetitive processes and deployment of new digital and physical services. However, library budget cuts urge the CBA to keep improving its understanding and prioritization of the information collected for budget decision making. As a consequence, the proposal to implement a holistic approach to support decision-making in the CBA academic library was presented to its authorities. The project started in 2010 and is to be finalized by the end of 2015.

A.3.2 First quadrant: internal perspective of the library system

This section documents the experience of applying TDABC to the four main traditional library functions performed in the CBA: acquisition, cataloging, circulation and document delivery. TDABC is a costing approach developed by Robert S. Kaplan and Steven R. Anderson in 2004 that requires only two parameters: 1) the unit cost of supplying resource capacity; and 2) an estimated time required to perform an activity (Kaplan & Anderson, 2007b). To calculate the activity costs through a TDABC model, this study followed the six steps presented by Patricia Everaert, Werner Bruggeman, Gerrit Sarens, Steven R. Anderson and Yves Levant (2008), which are described in detail by Lorena Siguenza-Guzman, Alexandra Van den Abbeele, Joos Vandewalle, Henri Verhaaren, and Dirk Cattrysse (2013)¹⁹. As a result, twelve processes were identified and analyzed (Table A.1).

¹⁹ Updated reference: Siguenza Guzman, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., Cattrysse, D. (2014). Using Time-Driven Activity-Based Costing to support Library Management Decisions: A Case Study for Lending and Returning Processes. *Library Quarterly*, 84 (1), 76-98.

Table A.1: Processes analyzed using TDABC

Area	Process
Acquisition	Books acquisition
	Journals acquisition
Cataloging	Original cataloging
	Copy cataloging
	Cataloging Closed-Stack Items
Circulation	Lending items
	Returning items
	Reference
Document Delivery	Requesting closed stack items
	ILL Outgoing Request
	ILL Incoming Request – digital items
	ILL Incoming Request – printed items

The application of TDABC to the CBA showed important benefits such as: 1) *Disaggregated values per activity*. Thanks to the TDABC implementation, many relevant findings on the process costs were unveiled. For instance, it was detected that scanning papers was a time consuming activity because the scanner was outdated (Pernot et al., 2007). Overdue fines consumes a significant part of the librarian’s time (5 times more costly than returning activities). A remediation strategy is to consider that fines could be paid by the annual university tuition fee. In addition, depending on the characteristics of an ILL request, “Searching” activities consumes approximately 75% of the process time (Pernot et al., 2007). This situation can be improved by simplifying the search process and outsourcing the activity to the requester. 2) *Allows comparison of different scenarios*. For instance, manually returning is almost 60% much more costly than the same activity performed using a self-check machine (Siguenza-Guzman, Van den Abbeele, et al., 2014b). In addition, TDABC showed that copy cataloging is 30% less time-consuming and thus less costly than original cataloging. 3) *Allows the justification of decisions and choices*. For instance, hiring students to work in activities such as photocopying, shelving, and scanning can reduce up to 25% of costs and, in turn, allows librarians to perform other specialized activities. Furthermore, librarians can propose the development of new services based on their responsibilities and time availability.

Nevertheless, a number of challenges were found during the TDABC implementation. For instance: 1) *Time*: data collection on the duration of activities took significant time, as the measuring was gathered by direct observation. Data were collected multiple times using a stopwatch during several days in the first semester of 2010, and then validated through an additional data collection in the second semester. Moreover, documenting the activity flows required considerable time. Two rounds of interviews were conducted with library managers and staff in order to identify the activities, resources and responsible. This step was improved by combining MS Visio and MS Excel to store, analyze and create graphical representations of the activity flows. This combination allowed validating the collected information straightforwardly because librarians could easily understand the sequences and their responsibility in each process. However, a dedicated software tool to perform TDABC analysis is strongly recommended in order to keep the flows updated and consequently to facilitate long-term maintenance. 2) *Feeling controlled*: some staff members felt uncomfortable being observed while working. This discomfort caused some resistance and consequently delayed the data collection. A right communication as well as the involvement and commitment of the managers and staff can increase the level of acceptance. In addition, library managers and TDABC team should explain the purpose of measurement, importance of the model, activities to perform, and implications of the results. They should clearly state that the activities and profiles are measured, not the names of individuals.

A.3.3 *Second quadrant: external perspective of the library system*

In the CBA case study, the LibQUAL+® survey was utilized to assess library service quality from an external perspective. LibQUAL+® is a set of services based on Web surveys that allows requesting, tracking and understanding users' perceptions of the library service quality (Association of Research Libraries, 2013). Three dimensions are measured in this survey: Affect of Service, Information Control, and Library as a Place. In 2008, KU Leuven was the first Belgian institution that used LibQUAL+® for assessing its services (University Library Services, 2009). This survey was a full version consisting of 45 questions, each requiring a response on a nine-point scale for current perceptions, as well as, minimum and desired expectations. Survey results showed that users were generally very satisfied with the library services and collection. The Library as a Place dimension was the best scored as its punctuation was slightly below the desired level. The former and Information Control obtained the same score; however, users' expectations for Information Control were 15% higher. Users were also satisfied with the Affect of Service dimension, especially researchers and academic staff.

In 2012 KU Leuven chose the LibQUAL+® Lite version including 23 questions in total in order to increase participation ratios. As a consequence, the total number of respondents increased 47% compared with the previous survey. In general, CBA performed very well in the dimensions Library as a Place and Affect of Service, but less in the dimension Information Control. In comparison with the already overall high score of the previous survey, the CBA was rated 4% higher. The Library as a Place dimension was the highest perceived score, even slightly higher than in 2008. Therefore, the importance of the Library as a Place is evidently still a concern, especially for students demanding more areas for individual and group work. In the Affect of Service dimension, CBA ranked 4% higher than the first survey. Survey results showed a positive effect on perceived service value. In addition, users' expectations (minimum and desired scales) are even higher in comparison with LibQUAL+® 2008, placing considerable value on a courteous and knowledgeable staff. Ultimately, Information Control was the relatively weak dimension. CBA ranked 3% lower than the previous survey. One reason was that a new search platform to access the collection was implemented at institutional level and a stabilization period was performed. Another reason was that while students still consider the physical collection very relevant and perceived that was not updated; researchers and academic staff are expecting and demanding more number of e-journals. After analyzing the survey results, several actions are being taken as shown in Table A.2 (Nassen, 2013; University Library Services, 2012).

The CBA had to deal with several challenges during the deployment period: 1) *Data Preparation*, due to specific factors such as language definition and great variety of population. Although LibQUAL+® provides the standard questionnaire in different languages, it was necessary to make some specific changes that had to be coordinated together with the Association of Research Libraries (Vandoolaeghe, 2013). In addition, KU Leuven chose to apply the survey in two languages: Dutch as primary language and English. This decision had important consequences such as the need of integrating the two results during the data processing. On the other hand, gathering the population data for each user group was not an easy task. In the case of students, for instance, it was important to distinguish the year that the student was being trained and not the year that the student was registered. Similarly, determining the number of PhDs by discipline was not a simple exercise because of the multidisciplinary groups. 2) *Granularity*, as no specific results for branch libraries and disciplines are provided by standard reports. LibQUAL+® survey produces standard reports in which the measurement is carried out in its overall performance and user groups (e.g. students, PhD's, faculties). This standard report also provides no direct insight into the library performance compared to other libraries branches where the LibQUAL+® survey was performed. As consequence, an online tool developed by Datimpact was appealed to analyze each sub-library (University Library Services, 2012). 3) *Participation rates*. Although LibQUAL+® Lite improved response rates and reduced respondent burden, there was still the perception that the survey was very long (Vandoolaeghe, 2013). Other strategies to stimulate the users' response were to offer incentives such as electronic devices and movie tickets, as well as to send a reminder email two weeks before the close of the survey. In addition, an input field for free text gave users the

opportunity to submit comments regarding their concerns and to express suggestions for future improvements. Eventually, the University involvement was crucial to obtain good results and libraries were totally aware of the importance of this analysis.

Table A.2: Action points after LibQUAL+® 2012

Domains	Actions points
Library as a place	Improvement of four group work areas and learning center facilities with high tech equipment such as smart boards, flat screens, and furniture.
Affect of service	Continuation of the customer service training programs, including student library employees. Development of long-term cooperation projects to exchange expertise in other library functions such as acquisition and cataloging. Cluster librarians that provided support on a number of related subjects became "Information Specialists". It meant going to the field to have a close contact with the researchers in order to know their expectations and needs.
Information control	Enhancing of the new search platform to provide easy access to all materials of its library collection. Improvement opportunities of remote access to online library resources and information services

A.3.4Third quadrant: external perspective of use

In the third quadrant, the theoretical framework evaluates the usefulness of the library collection. To do so, Siguenza-Guzman et al. propose to combine citation analysis, vendor-supplied statistics and citation databases such as PubMed, Scopus, Web of Science, and Google Scholar to gain extensive knowledge about the value of the library collection. At the CBA, an ambitious project that combines the three methodologies is currently being performed. The aim of this project is to have a deep insight of the local use of the collection, with especial interest on the e-journal's availability. Thus, more than 1,200 PhD theses submitted over a six-year period (2005-2010) are being analyzed. These theses correspond to researches conducted in 13 departments of Science, Engineering and Agriculture of the KU Leuven. As a result about 235,000 references are being collected and evaluated. The results will allow to personalize reports based on the library requirements such as journals cited per department, workgroup, and advisor. The project is expected to be concluded by the end of June 2013.

The study first collects in a database all references cited in each PhD theses. In parallel, a second database is created gathering information about the publishing patterns of PhD students. This second database allows determining the most attractive journals where departments choose to publish, as well as verifying whether these journals correlate with the citations used as reference. A third database is used to collect the vendor-supplied statistics of all journals downloaded during the period 2005-2010. These electronic journal usage data are received from COUNTER-compliant publishers as part of the subscription contract. The Counting Online Usage of NeTworked Electronic Resources (COUNTER) standards are an internationally accepted initiative that facilitates the recording and exchange of online usage data in a consistent, credible and compatible manner (COUNTER, 2013). This third database verifies the correlation among the citations patterns, publishing patterns and journals downloaded. The information collected in these three databases is then used to test an additional correlation with the 5-year Impact Factor produced by Thomson ISI Web of Knowledge. Finally, as a result of the previous analysis, a list of journals is created and classified according Bradford's Law in order to determine the core collection of the library.

The implementation of this analysis has faced several challenges, such as: 1) *Time*. To manually analyze a thesis, it is necessary an average of 2.5 hours to both collect the information and incorporate them in the different databases. However, in order to facilitate long-term maintenance, process automation is necessary. 2) *Abbreviations*, there is no defined standard for journals' abbreviations and acronyms, thus collecting journals' information is not always straight forward. For instance, the ISO Abbreviation for the Journal of the American Chemical Society is J. Am. Chem. Soc.; the JCR Abbreviation is J AM CHEM SOC, while its acronym is JACS. Proc. IEEE is the ISO abbreviation for The Proceedings of the IEEE and its JCR abbreviation is P IEEE. Therefore, a certain expertise is necessary to differentiate the different abbreviations and acronyms of journals that PhD students cited as reference. 3) *Data management*, although the project is using Excel sheets as main platform, there is the need of dedicated software to collect the large amount of information, as well as to evaluate the results.

A.3.5 Fourth quadrant: internal perspective of use

The final quadrant measures users' interaction with the system. Siguenza-Guzman et al. suggest the use of transaction log analysis to monitor users' behavior in a digital environment. To date, at the CBA no prior studies have assessed users' behavior. Therefore, a project to analyze transaction logs is expected to start in July 2013 and to be concluded by the end of June 2014. Examples of challenges that this project will face include: 1) User privacy. Privacy of personally identifiable user information is of concern during the bibliomining process (Nicholson, 2006b). Several solutions have been proposed in literature such as to encode the user identification in the data warehouse by replacing the user ID with a code. Another option is to create a demographic surrogate to replace personal information about the user through a set of demographic values (e.g. age, sex, education). 2) Identifiability of IP Addresses. Because CBA is part of a University system, it is required to carefully define the range of IP addresses to be monitored.

A.4 Conclusion

To holistically analyze a library, several parameters must be considered including both the library functions (collection and services), as well as stakeholders' perception (internal and external). In this paper, the implementation of a set of methodologies and measurement tools has been described. We conclude that the model proposed by Siguenza-Guzman et al. is a simple and powerful structure for grouping the library information prior a decision making. By documenting the initial stages of implementation, this paper provides preliminary experiences supporting the practical validity of the proposed holistic approach in order to enable a budgeting decision-making process. There are, however, important considerations to be borne in mind such as the time required to implement the complete approach, as well as the need of dedicated systems to automate the different quadrants.

References

- Association of Research Libraries. (2013). LibQUAL+® Web site: Retrieved February 18, 2013, from <http://www.libqual.org/home>
- COUNTER. (2013, February). About COUNTER. Retrieved from <http://www.projectcounter.org/>
- Editors of the American Heritage Dictionaries. (2011). The American Heritage Dictionary entry: holistic. The American Heritage Dictionary. Dictionary. Retrieved November 8, 2012, from <http://www.ahdictionary.com/word/search.html?q=holistic&submit.x=56&submit.y=19>
- Ernst, D. J., & Segall, P. (1995). Information Resources and Institutional Effectiveness: The Need for a Holistic Approach to Planning and Budgeting. *CAUSE/EFFECT*, 18(1), 11–13.
- Everaert, P., Bruggeman, W., Sarens, G., Anderson, S. R., & Levant, Y. (2008). Cost modeling in logistics using time-driven ABC: Experiences from a wholesaler. *International Journal of*

- Physical Distribution & Logistics Management, 38(3), 172-191. doi:10.1108/09600030810866977
- Kaplan, R. S., & Anderson, S. R. (2007). Time-driven activity-based costing: a simpler and more powerful path to higher profits. Harvard Business Press.
- Nassen, C. (2013, February 20). Experiences on conducting LibQUAL+® survey in CBA.
- Nicholson, S. (2004). A conceptual framework for the holistic measurement and cumulative evaluation of library services. *Journal of Documentation*, 60(2), 164-182. doi:10.1108/00220410410522043
- Nicholson, S. (2006). The basis for bibliomining: Frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services. *Information Processing & Management*, 42(3), 785-804. doi:10.1016/j.ipm.2005.05.008
- Pernot, E., Roodhooft, F., & Van den Abbeele, A. (2007). Time-Driven Activity-Based Costing for Inter-Library Services: A Case Study in a University. *The Journal of Academic Librarianship*, 33(5), 551-560. doi:10.1016/j.jalib.2007.06.001
- Siguenza-Guzman, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2013). Using Time-Driven Activity-Based Costing to Support Library Management Decisions: A Case Study for Lending and Returning Processes. submitted to *Library Quarterly*.
- Siguenza-Guzman, L., Van Den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2013). A holistic analysis as a key role for supporting academic libraries in the resource allocation process. submitted to *Library Quarterly*.
- University Library Services. (2009, July 7). Library users satisfied but the library will work even harder. University Library KU Leuven. Retrieved February 18, 2013, from <http://libkuleuven.wordpress.com/2009/07/07/library-users-satisfied-but-the-library-will-work-even-harder/>
- University Library Services. (2012, December 7). LibQUAL+ 2012. University Library. Retrieved February 19, 2013, from <http://bib.kuleuven.be/ub/over/tevredenheidsenquete/libqual%202012>
- Vandoolaeghe, F. (2013, February 19). Experiences on conducting LibQUAL+® survey at institutional level.
- Wright, S., & White, L. S. (2007). Library Assessment: SPEC Kit 303. Association of Research Libraries, 14.

Appendix B: Time-Driven Activity-Based Costing Systems for Cataloging Processes

Siguenza-Guzman, L., Van den Abbeele, A., & Cattryse, D. (2014). Time-Driven Activity-Based Costing Systems for Cataloging Processes: A Case Study. The Liber Quarterly, 23 (3), 160-186.

This appendix complements Chapter 4 by describing a case study of the TDABC implementation in one of the most important library processes as is the cataloging. In particular, original and copy cataloging are analyzed through a case study to demonstrate the applicability and usefulness of TDABC to perform cost analysis of cataloging processes. The appendix starts by briefly outlining the theoretical background of costing systems. Then, the different steps involved in implementing TDABC in original and copy cataloging are explained. Next, similarities and differences found in the original and copy cataloging regarding time and cost per activity are discussed. In addition, a number of recommendations where process improvements were unveiled are also incorporated. Eventually, benefits encountered when implementing the TDABC model in an academic library in Belgium are described.

Apart from typographical adjustments, the content of this appendix is identical to the content of the published paper quoted above; where necessary, additional information or remarks are added in footnotes. The layout is adapted for consistency throughout this dissertation. Some redundancy with other chapters is unavoidable as an academic article needs its own introductory sections. This, however, entails the advantage that the chapter can be read separately.

Abstract

TDABC is a relatively new costing management technique, initially developed for manufacturing processes, which is gaining attention in libraries. This is because TDABC is a fast and simple method that only requires two parameters, an estimation of time required to perform an activity and the unit cost per time of supplying capacity. A few case studies have been documented with regard to TDABC in libraries; all of them being oriented to analyze specific library activities such as inter-library loan, acquisition and circulation processes. The primary focus of this paper is to describe TDABC implementation in one of the most important library processes, namely cataloging. In particular, original and copy cataloging are analyzed through a case study to demonstrate the applicability and usefulness of TDABC to perform cost analysis of cataloging processes.

Contributions of the first author

The first author's contributions are: the literature study on costing systems, with emphasis on TDABC, TDABC implementation in original and copy cataloging, description of TDABC benefits in cataloging processes and conclusions.

B.1 Introduction

In the current economic situation, characterized by periodic shortages and limited budget resources, libraries are in search of methods to improve their process efficiency and provide high-quality services at lower costs (ACRL Research Planning and Review Committee, 2010; Cottrell, 2012). In order to improve their performance, library managers need to consider cost reduction and the inclusion of service costs in their decisions (Hoozée, Vermeire, & Bruggeman, 2012). Furthermore, they should strive to identify improvement opportunities and to eliminate costs related to non-value adding activities (Ellis-Newman, Izan, & Robinson, 1996). In order to do this, library managers must keep their activities, resources and costs under control, relying on valid information about activity costs, resource capacity and their performance (Stouthuysen, Swiggers, Reheul, & Roodhooft, 2010).

In service institutions such as libraries, several methodologies for costing analysis have been used for decades, whereas the Time-Driven Activity-Based Costing (TDABC) is one of the most recent approaches. TDABC is a cost management technique developed by Robert S. Kaplan and Steven R. Anderson to overcome the difficulties presented by previous costing systems (Kaplan & Anderson, 2007). By implementing TDABC in libraries, key benefits are expected, including the possibility of benchmarking different scenarios; identifying non-value added activities; and justifying decisions and choices for staff recruitment, training and new service development (Siguenza-Guzman et al., 2013a).

Although some research has been carried out with respect to TDABC in libraries, these studies have been focused on specific library activities such as the inter-library loan (ILL), acquisition, and circulation (Siguenza-Guzman et al., 2013b). More research is still required to determine whether TDABC can be useful in other library services. Therefore, the aim of the paper is to provide more detailed insight on implementing TDABC for library cataloging processes. We focus on this unit because cataloging is considered for many libraries, to be one of their most expensive processes, especially original cataloging (Manaf & Rahman, 2006). The remainder of this paper is organized as follows. Firstly, a brief outline of the theoretical background of costing systems is provided (§ 5.2). Secondly, the different steps involved in implementing TDABC in original and copy cataloging are explained (§ 5.3). Thirdly, we discuss the similarities and differences found in the original and copy cataloging regarding time and cost per activity (§ 5.4). In addition, a number of recommendations where process improvements were unveiled are included in this section. Fourthly, the benefits encountered when implementing the TDABC model in an academic library in Belgium are described (§ 5.5). We end this chapter with a brief conclusion (§ 5.6).

B.2 Theoretical background: Costing systems

B.2.1 Traditional Costing Systems

Several library cost analysis studies have been performed since the 1970's (Roberts, 2003). However, these studies were mainly treated as technical rather than organizational or managerial innovations (Kont, 2011). Jennifer Ellis-Newman, Haji Izan, and Peter Robinson (1996) report that the majority of prior studies on library costs were undertaken in the United States, utilizing cost allocation models more compatible with traditional costing methods. Traditionally, the total product cost consists of direct costs, such as the cost of materials and direct labor, and a percentage of overheads as indirect costs (Siguenza-Guzman, Van den Abbeele, et al., 2013). The latter includes training, marketing, and infrastructure, among others. Traditional costing systems are adequate when indirect expenses are low and product variety is limited. However, in an environment with a broad range of products and enhanced services, such as a library, indirect costs have become substantially more complex than direct costs (Siguenza-Guzman, Van den Abbeele, et al., 2013). This situation renders traditional methods inadequate; not only for estimating the effect of strategic

decisions, but also for providing crucial information to library managers (Ellis-Newman & Robinson, 1998; Kaplan & Cooper, 1998).

B.2.2 Activity-Based Costing Systems

Activity-Based Costing (ABC) is an advanced costing calculation that seeks to remedy the limitations of traditional methods (Kaplan & Cooper, 1998). ABC, promoted by Robin Cooper and Robert S. Kaplan in the mid-80s, first accumulates indirect costs for each activity and then assigns the activity costs to the services causing that activity (Cooper & Kaplan, 1988; Ellis-Newman & Robinson, 1998). ABC has proven to be a valuable tool for libraries through its implementation in several case studies. For instance, Ellis-Newman et al. (1996) examined the application of ABC in academic libraries of two Western Australian universities. This study illustrates how activity costing helps managers to differentiate key activities from others that do not add value. In an additional study, Jennifer Ellis-Newman and Peter Robinson (1998) discuss the benefits of ABC for library managers and the steps involved in implementing ABC in an academic library. The authors show how traditional costing systems are unable to explain the relationship between costs and the underlying activities. Furthermore, Jennifer Ellis-Newman (2003) demonstrates the type of information that an ABC system provides to assist decision-making with a case study in the user services area of an Australian academic library. In turn, Steve H. Ching, Maria W. Leung, Margaret Fidow and Ken L. Huang (2008) employ ABC to examine the Super e-Book Consortium in Taiwan and Hong Kong. The study finds cost drivers of consortia business operations, and identifies the key consortium activities and their relevant costs. Moreover, Andrew Goddard and Kean Ooi (1998) examined the development of ABC through a case study applied to library services at the University of Southampton. The authors present ABC as an option to overcome some of the problems of overhead allocations. Despite the benefits of ABC, they also describe significant problems with its practical application such as the amount of resources and time required for its development and maintenance. Eventually, Denise D. Novak, Afeworki Paulus and Gloriana St. Clair (2011) describe how a medium-sized university library implemented ABC and other decision-making strategies to make budgetary cuts and thereby, redirecting library services.

Although a relatively extensive stream of literature finds that ABC systems provide interesting advantages for decision making in libraries, ABC also has its limitations. Kaplan and Anderson (2004, 2007) note that ABC is difficult and costly to implement and maintain, especially when the current accounting system does not support the collection of ABC information. Data collection is time consuming and costly because of the need to interview and survey the library staff to estimate the percentage of time spent on each activity. While it works well in small organizations and limited activities, it becomes problematic to scale up to larger organizations (Hoozée et al., 2012). Managers also question the accuracy of the system since cost assignments are based on individuals' subjective information on how they spend their time. Furthermore, staff resistance could arise, as employees might feel threatened by the suggestion that their work should be improved. As a consequence, ABC systems tend to be outdated, and in some cases are abandoned and substituted by less demanding approaches such as Time-Driven Activity-Based Costing (Wegmann & Nozile, 2009; Yilmaz, 2008).

B.2.3 Time-Driven Activity-Based Costing Systems

TDABC is a useful cost management technique developed by Kaplan and Anderson in 2004 to overcome the difficulties presented by previous costing systems (Kaplan & Anderson, 2004). TDABC assigns resource costs directly to cost objects using a fast and simple framework that only requires the unit cost of supplying resource capacity, and an estimation of the time duration of an activity (Kaplan & Anderson, 2007). Unlike the percentages that employees subjectively estimate for an ABC model, the time duration in a TDABC model can be readily observed and validated (Kaplan & Anderson, 2007). For each activity, "costing equations" are calculated and computed by time equations, which are the sum of individual activity times (Yilmaz, 2008). Through the use of time equations, TDABC allows incorporation of variation in the time demands made by different types of transactions, and consequently the representation of all possible combinations of activities

that a process performs (Kaplan & Anderson, 2007). Five main TDABC advantages, highlighted by Lorena Siguenza-Guzman, Alexandra Van den Abbeele, Joos Vandewalle, Henri Verhaaren, and Dirk Cattrysse (2013), are its simplicity to build accurate models and improve the understanding of the different processes; the opportunity of modelling complex operations thanks to the use of multiple drivers; a good estimation of resource consumption and capacity utilization; its fast maintenance compared to ABC models; and the possibility of using TDABC in a predictive manner.

TDABC has been carried out in specific library activities such as inter-library loan (ILL), acquisition and circulation. For instance, the first case study by Eli Pernot, Filip Roothoof and Alexandra Van den Abbeele (2007) uses TDABC to calculate ILL costs and describes TDABC as a useful technique to reduce ILL resource costs and to renegotiate ILL service prices based on more accurate costs. The authors conclude that TDABC is very suited to cope with increasing cost pressures, and its findings can contribute to improve library services at lower costs. A second case study presented by Kristof Stouthuysen, Michael Swiggers, Anne-Mie Reheul, and Filip Roodhooft (2010) describes the use of TDABC for a library acquisition process. The authors state that TDABC provides library managers with a better insight into cost drivers, visualizes the acquisition process efficiencies and capacity utilization, and leads to potential cost efficiencies. As an illustration, they consider that 50% of some costs could be saved if administrative assistants get involved in the acquisition process instead of being performed by the head of department, provided that they are capable of doing these tasks. In addition, the authors demonstrate that TDABC can be updated rapidly and inexpensively to changes. Due to this flexibility, they consider that TDABC can be applied to other processes such as cataloging or digitalized activities, with significant benefits. The latest case study by Siguenza-Guzman, Van den Abbeele, Vandewalle, Verhaaren and Cattrysse (2014) uses TDABC to analyze lending and returning processes. The authors provide several important insights for a successful implementation of TDABC in libraries such as: 1) collecting the time duration of the activities through direct observation to improve the level of accuracy; 2) using graphical representations of activity flows to validate the collected information straightforwardly; and 3) showing that clarifying the measurement purpose is crucial in improving the level of acceptance, and achieving the desired commitment with staff. They conclude that the TDABC implementation is worthwhile since it leads to a more accurate cost and process analysis for supporting decision-making.

B.3 TDABC in Cataloging Processes

The data used for this research was collected at the Arenberg Campus Library²⁰ (Campusbibliotheek Arenberg, hence the abbreviation CBA) of KU Leuven in Belgium. CBA offers information sources on subjects of the exact sciences, engineering, architecture, kinesiology, and rehabilitation sciences (Campus Bibliotheek Arenberg, 2013). Its services are handled by approximately 20.5 full-time equivalent employees (FTE). In this case study, we focus on describing the application of TDABC in two types of cataloging activities: original and copy cataloging. The former refers to creating a new bibliographic record from scratch, while the latter to adapting a pre-existing record to the characteristics of the item in hand (Reitz, 2004).

For this case study, qualitative interview data with quantitative data analysis were combined following the six steps presented by Patricia Everaert, Werner Bruggeman, Gerrit Sarens, Steven R. Anderson and Levant (2008) to calculate the cost of activities through the TDABC model. These steps illustrated in Table B.1 are described in detail by Siguenza-Guzman et al. (2014).

To identify resource groups involved in cataloging activities as required in Step 1, multiple interviews were conducted. Initial interviews started through brief discussions with the library manager, and then moved to a more detailed level with the library staff. For each activity, a final

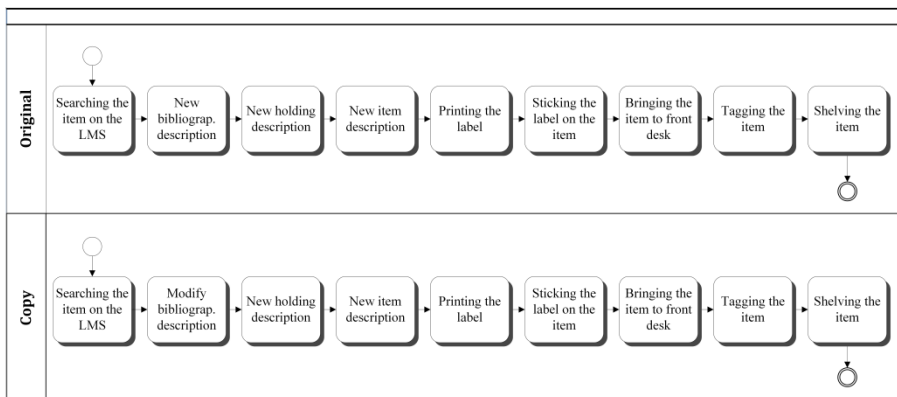
²⁰ <http://bib.kuleuven.be/cba/>

interview was performed in order to validate specific details about the different sub-activities. During the interviews, key activities involved in each process were described in detail by employees in charge. The information was used to build flow charts of activity sequences. As Siguenza-Guzman et al. (2014) indicate, flowcharts allow a good overview of the different activities performed in a process, to identify additional expenditures such as computer maintenance and software licenses, and afterwards to validate the activities in an optimal and simple manner. Figure B.1 shows the activity flow of original and copy cataloging respectively.

Table B.1: Time-Driven Activity-Based Costing steps (Everaert et al. 2008)

Step	Description
1	Identification of resource groups
2	Estimation of the total cost of each resource group
3	Estimation of the practical capacity of each resource group
4	Calculation of the unit cost of each resource group
5	Estimation of the standard time duration of each activity
6	Multiplying the unit cost of each resource group by the time duration per activity

Cataloging Processes



Note. – LMS = Library Management System.

Figure B.1: Cataloging Processes: a) Original Cataloging; b) Copy Cataloging

The total cost of each resource group required in Step 2 was provided by the accountant and library manager via the Library Management System (LMS). The costs were classified into direct and indirect costs. Direct costs included salaries of staff and student library employees – SLE (i.e. students hired to perform secondary activities), equipment, and technology. Conversely, examples of indirect costs included stationery, electricity, support, telephone, training, and other items used to perform an activity (Vazakidis & Karagiannis, 2009).

The salaries of catalogers were calculated based on the average salary earned by employees responsible for cataloging. According to the Chief Librarian, the total number of personnel assigned to cataloging represents 1 full-time equivalent (FTE). The 1 FTE consisted of six people, each dedicating different amounts of their time to cataloging processes. This corresponded to about € 59,000 on a yearly basis. LMS costs covered the annual integrated library system and supporting

software license fees attributed for the CBA. The Library Management System includes functionalities for acquisition, cataloging, circulation and reporting. This integrated library solution was acquired for the entire library which consisted of the Central Library, three Campus Libraries and several Faculty Libraries for the Humanities and Social Sciences group. The LMS costs amounted to € 17,000 on a yearly base. The annual computer maintenance costs for specific tasks such as reparation, maintenance, cleaning and depreciation of a PC in the cataloging processes equaled € 5,000. RFID maintenance costs refer to the costs associated with the maintenance, repair and inspection of the RFID system and yearly costs corresponded to about € 17,000.

In the case of indirect costs, the library accountant estimated that about € 195,000 was annually spent on general overhead (GO) costs. GO costs included management, secretary, accounting, training, staff meetings and stationery material. Other indirect values such as electricity, telephone, heating and transportation were not accounted as part of general overhead costs since they were paid by the University and not charged to the library (Ellis-Newman & Robinson 1998). In order to calculate the overhead costs attributed to cataloging activities, general overhead cost was divided by the total number of FTE working at the entire library. This resulted in a yearly overhead of approximately € 9,500 per FTE. An overview of the total cost of each resource group can be seen in Table B.2.

Table B.2: Total cost of each resource group

Resource Group	Cost (€) per year
Catalogers	€ 59,000
Library Management System	€ 17,000
Computer Maintenance	€ 5,000
RFID Maintenance	€ 17,000
General Overhead	€ 9,500

Then, the practical capacity estimation of each resource group was calculated in Step 3. According to Kaplan and Anderson (2007, pp. 52–53), practical time capacity can be estimated in two different ways: 1) assuming an 80% of theoretical time capacity for people due to breaks, arrival and departure, training, meetings, and chitchats; and 85% for machines due to maintenance, repair, and scheduling fluctuations. 2) Calculating the real values according to the library situation, for example, available working hours, excluding holidays, meetings and training hours. In order to simplify the study, the first option was selected. For staff capacity, 38 hours per week were accounted as theoretical time capacity. It means 30.4 hours per week for practical capacity (staff practical capacity = 38 hours * 80%). Assuming fifty-two weeks per year, the practical capacity of a cataloger is equal to 94,848 minutes per year $\left(30.4 \frac{\text{hours}}{\text{week}} \times 52 \frac{\text{weeks}}{\text{year}} \times 60 \frac{\text{minutes}}{\text{hour}}\right)$. In the case of machines, the theoretical time capacity was set equal to the time in which catalogers were available that is again 38 hours per week. Thus, the practical capacity for machines is 32.3 hours per week (machines practical capacity = 38 hours * 85%), and 100,776 minutes per year $\left(32.3 \frac{\text{hours}}{\text{week}} \times 52 \frac{\text{weeks}}{\text{year}} \times 60 \frac{\text{minutes}}{\text{hour}}\right)$.

Once the practical capacity was obtained, the cost per unit time was calculated in Step 4 by dividing the total cost of a resource (Step 2) by the practical time capacity (Step 3). An overview of the resulting costs involved in this analysis is shown in Table B.3.

$$\text{Cost per unit time} = \frac{\text{total cost of the resource}}{\text{practical capacity}}$$

Table B.3: Costs involved in the analysis

Resource Group	Cost per minute (€/min)
Catalogers	0.62
Library Management System	0.17
Computer Maintenance	0.05
RFID Maintenance	0.17
General Overhead	0.10

For the fifth step, Kaplan and Anderson (2007, p. 30) recommend that the time required to perform an activity should be estimated based on standard times rather than actual times, since these times might reflect random variation, individual employee variation and nonrecurring factors. Moreover, the authors argue that precision is not critical and that a rough accuracy is sufficient because gross inaccuracies will be revealed either in unexpected surpluses or shortages of committed resources. This level of accuracy can be obtained by multiple methods such as direct observation, interviews, process maps or leveraging time estimates from elsewhere in the institution (Kaplan & Anderson, 2007, p. 26). In our case study, the standard time to perform an activity was gathered in the academic period 2010 – 2011 through direct observation as recommended by Siguenza-Guzman et al. (2014). Observations were made multiple times using a stopwatch during several days at different hours in order to avoid possible biases (Siguenza-Guzman et al., 2014). Finally, based on their average values, an additional interview was performed in order to validate the data collection. Next, time equations were constructed for each activity. Time equations are the sum of individual activity times, which are represented with the following expression (Kaplan & Anderson, 2007):

$$\begin{aligned} \text{Time required to perform an activity} \\ = (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \dots + \beta_i X_i) \end{aligned}$$

With:

β_0 = The standard time to perform the basic activity (e.g. 2 minutes)

β_i = The estimated time for the incremental activity i (e.g. time required for a librarian to enter an item in the cataloging system = 0.5 minutes)

X_i = The quantity of incremental activity i (e.g. items per batch = 1, 2...)

B.3.1 Original Cataloging

The process, as shown in Figure B.1a, starts by searching the item in hand on the LMS in order to verify whether a similar record is already present in the database. This searching takes on average of 57s. If the item and record do not appear to match, the cataloger creates a new record which includes the bibliographic description, requiring 306s (or 5min and 6s). A bibliographic description is the standardized description of an item including: title, edition, material specific details, details of publication, standard number, etc. (Reitz, 2004). The cataloger then creates a new holding description, which is usually the information concerning the location of an item, taking 32s; and finally a new item description in 15s. Item description indicates item type, volume number, barcode, and loan rules. Once the new record is processed and stored in the database, the cataloger prints the corresponding label (30s), and sticks on the item in 28s. Afterwards, the cataloger brings the item to the front desk (30s) in order to tag the item (40s). An individual tag costs €0.30 including VAT. Eventually, the item is placed on the corresponding shelf or stack, requiring 168s (or 2min and 48s).

B.3.2 Copy Cataloging

In contrast to the above process as shown in Figure B.1b, the cataloger does find a record that appears to match with the item in hand. The cataloger requires 74s (or 1min and 14s) to validate and modify the bibliographic description. Next, the cataloger creates a new holding and item description, taking 32s and 15s respectively. Labelling and shelving are the same as the original cataloging process.

As the only difference between original and copy cataloging is new versus modifying bibliographic description, we can create only one time equation by adding dummy variables in the equation. A dummy is a variable that takes the value 1 or 0 if a certain condition is true or false respectively. The resulting equation is as follows:

Cataloging

= Searching + New_{Bibdescription} {if original_cataloging}
+ Mod_{Bibdescription} {if copy_cataloging} + New_{holding}
+ New_{item} + Print_{label} + Stick_{label} + Bring_{item} + Tag_{item}
+ Shelf_{item}

Cataloging

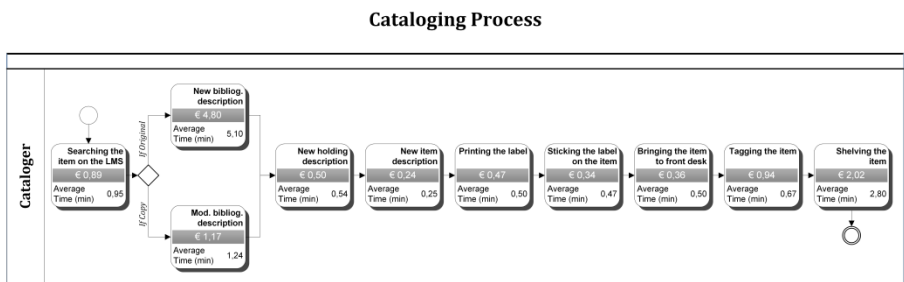
= 0.95 + 5.10 {if original_cataloging} + 1.24 {if copy_cataloging}
+ 0.54 + 0.25 + 0.50 + 0.47 + 0.50 + 0.67 + 2.80

Once the estimated time per activity and the unit cost of each resource group is calculated, costs are assigned to cost objects by multiplying the unit cost per time of resources by the estimated time required to perform the activity. It is represented by the following mathematical expression:

*Cost of an individual activity = time required to perform an activity * resources cost*

Figure B.2 shows the resulting activity flow of the cataloging process including average times and costs per sub-activity. Original and copy cataloging are integrated in the same graphical representation by using decision diamonds. Finally, the total cost of the process is calculated by summing up all activity costs and can be represented by:

Total Cost of a Process = ∑ costs of individual activities



Note. – LMS = Library Management System.

Figure B.2:: Resulting activity flow of the Cataloging Process

The total costs and time incurred in the original and copy cataloging processes are presented in Table B.4. The resulting table is divided vertically into six columns and horizontally by standard and optional activities to separate the activities influenced by dummy variables. The first column lists the activities identified in the cataloging process; the second column shows the average time per activity. The third column specifies the accumulated costs per minute of resources involved in each activity; the fourth column calculates the resulting cost incurred in the activity. The fifth column describes the condition under which two options are available to be selected: original and copy cataloging. Eventually, the sixth column indicates the resource groups involved per activity.

The subtotal of “standard activities” is the sum of costs included in both processes: original and copy cataloging. The average time in minutes for standard activities was 6.68 with an activity cost of €5.77. To calculate the total cost of the original cataloging process, standard activities were summed up with the optional activity, new bibliographic description. The resulting time in minutes was 11.78 with a cost of €10.57. On the other hand, to calculate the total cost of the copy cataloging process, standard activities were summed up with the optional activity, modify bibliographic description. The results show that the average time for copy cataloging was 7.92 minutes and its cost was €6.94 per title.

B.4 Original and Copy Cataloging

Because of budgetary constraints and technological changes, cataloging units have significantly influenced the nature of cataloging work (Mitchell, Thompson, & Wu, 2010). In order to become more efficient, catalogers constantly search for new ways to increase bibliographic access without spending more money (Morris & Wool, 1999). That includes automation, outsourcing, lowered costs for traditional cataloging, and an increasing variety of information resources to control. With regard to lowering costs, a number of studies on cataloging costs have been reported in the literature. For instance, the Iowa State University (ISU) conducted a longitudinal time-cost study to investigate the impact of automation on cataloging costs (Morris, Hobert, Osmus, & Wool, 2000). Results showed that in 1987-1990, the average cost of original cataloging at ISU was \$34.13 and that copy cataloging was \$8.18 (Morris, 1992). In 1997-1998, original cataloging costs performed by faculty catalogers increased to \$75.43 and to \$58.72 when some original records were contributed by library assistants. Copy cataloging costs, in the same period, increased to \$35.82 when performed by faculty catalogers and to \$8.87 when performed by library assistants. Although costs of original and copy cataloging increased during the time, Dilys E. Morris, Collin B. Hobert, Lori Osmus, and Gregory Wool (2000) also highlighted the fact that the average cost of cataloging processes (i.e. copy, full original, minimal original and recataloging) declined from \$20.83 to \$16.25 between 1990/91 and 1997/98. Authors attributed this decrease to national collaborative efforts, technological development, and reengineering efforts that have improved costs effectiveness and quality process (Morris & Wool, 1999). The University of Oregon, following the previous research at ISU, conducted a benchmark analysis during autumn 1997 to determine time and costs for acquisitions, cataloging, and processing functions (Slight-Gibney, 1999). Results showed that the average cost for copy cataloging was \$9.23 per title, and that original cataloging was \$24.92. A third case study by Ellis-Newman and Robinson (1998) in an Australian academic library, reported the cost of library services utilizing ABC models. Results showed that the average cost of copy cataloging at Edith Cowan University was \$12.48 and original cataloging was \$54.39. Authors highlighted the importance of implementing activity-based costing systems in libraries to assign more accurate costs to services, to categorize costs, and to develop a price schedule for fee-based services.

In Table B.5, all cases cataloging costs vary library to library, even if the change in the rate of exchange and inflation are included in the three literature case studies. Michael D. Charbonneau (2005) points out that these numbers logically vary because they are based on locally produced data and operations, individual cataloging expertise, the type of material cataloged, and the cataloging tools and resources available. Differences in overhead costs can also explain some of the cost variations. For instance, cataloging overhead costs at ISU represent approximately 45% of the full costs, while in our case study about 10%. Unfortunately, not all studies provide detailed

Table B.4: Total cost of the cataloging process

Activity	Average Time (Minutes)	Cost per min (€/Minute)	Cost (€)	Condition	Resources
<i>Standard activities:</i>					
Searching the item on the LMS	.95	.94	.89		Cataloger + LMS + CM + GO
New holding description	.54	.94	.50		Cataloger + LMS + CM + GO
New item description	.25	.94	.24		Cataloger + LMS + CM + GO
Printing the label	.50	.94	.47		Cataloger + LMS + CM + GO
Sticking the label on the item	.47	.72	.34		Cataloger + GO
Bringing the item to front desk	.50	.72	.36		Cataloger + GO
Tagging the item	.67	1.41	.94		Cataloger + LMS + CM + GO + RM + Tag
Shelving the item	2.80	0.72	2.02		Cataloger + GO
Subtotal	6.68		5.77		
<i>Optional activities:</i>					
New bibliographic description	5.10	.94	4.80	if original_cataloging	Cataloger + LMS + CM + GO
Modify bibliographic description	1.24	.94	1.17	if copy_cataloging	Cataloger + LMS + CM + GO
Subtotal		
Total Original Cataloging	11.78		10.57		
Total Copy Cataloging	7.92		6.94		

Note. – LMS = Library Management System. GO = General Overhead. CM = Computer Maintenance. RM = RFID Maintenance..

Table B.5: Cataloging costs adjusted using inflation and exchange-rates

University	Period	Original Cataloging	Copy Cataloging	Inflation	Inflated Original Cataloging	Inflated Copy Cataloging	EUR Conversion	Original Cataloging (€)	Copy Cataloging (€)
Iowa State University (US)	1987-1990	USD \$34.13	USD \$8.18	68.66%	57.56	13.80	0.72	41.39	9.92
Iowa State University (US)*	1997-1998	USD \$75.43	USD\$35.82	37.69%	103.86	49.32	0.72	74.68	35.46
Iowa State University (US)†	1997-1998	USD \$58.72	USD \$8.87	37.69%	80.85	12.21	0.72	58.13	8.78
University of Oregon (US)	1997	USD \$24.92	USD \$9.23	39.91%	34.87	12.91	0.72	25.07	9.29
Edith Cowan University (AU)	1998	AUD\$54.39	AUD\$12.48	47.20%	80.06	18.37	0.72	59.41	13.63
KU Leuven (BE)	2010-2011	EUR€10.57	EUR €6.94	---	---	---	0.74	10.46	6.83

Note.- Inflation data from inflationdata.com (US, Dec 1987/1998 to Dec 2011), and rba.gov.au (AU, 1998 to 2011)

Exchange data from x-rates.com (USD and AUD to EUR in 2011)

* Cataloging costs when performed by a faculty cataloger

† Cataloging costs when performed by a library assistant

information on the overhead costs. In this case study, these costs include management, secretary, accounting, training, staff meetings and stationery material; however, other university costs such as electricity, telephone, heating and transportation are not accounted for since they are not charged to the library. At ISU overhead costs are organized into overhead centers: paid leave, automation and support services (Morris et al., 2000). Some other studies consider overhead costs as approximate since they are estimates or based on incomplete data (Slight-Gibney, 1999). Additional factors that vary among libraries are the library structure and process flow, for instance, pre-cataloging activities are included in the cataloging costs of one library (e.g. ISU), by the acquisitions department at another library, or even delegated to library assistants or students. In our case study, this activity is not included in the cataloging flow because it is performed by the person responsible for acquisitions. Finally, differences in the periods of analysis can also explain such cataloging costs variations, especially due to evolving factors that need to be taken into account. These factors include the increasing automation of cataloging activities, use of shared cataloging and authority records, decreasing staffing for cataloging, and the growing presence of new information formats (Morris & Wool, 1999).

Although cataloging costs data are not necessarily comparable among libraries (McCain & Shorten, 2002); “best practices” can be adopted in other libraries to influence their own workflows and to streamline their processes. This case study contributes to the cataloging cost study literature by providing an additional approach for calculating cataloging costs based on a fast and simple method as is TDABC. In this case study, the copy cataloging, as shown in Figure B.3, is approximately 34% less costly and 33% less time-consuming than original cataloging. Thanks to the TDABC’s ability to disaggregate costs per activity, it is possible to clearly analyze which activities demand more time, and thus lead to higher costs. For instance, the unique difference between original and copy cataloging is the creation or modification of a bibliographic description. Results shown in Figure B.4 indicate that adapting a bibliographic description from a pre-existing record is approximately 75% less costly than creating a new full bibliographic record. Based on these findings, it is concluded that is worthwhile to recommend librarians to adapt a pre-existing record instead of creating a new bibliographic record from scratch, as this will lead to significant costs and time reductions. Copy cataloging not only increases catalogers efficiency by eliminating duplication of effort, but also by reducing typographical errors caused in local libraries (Beall & Kafadar, 2004).

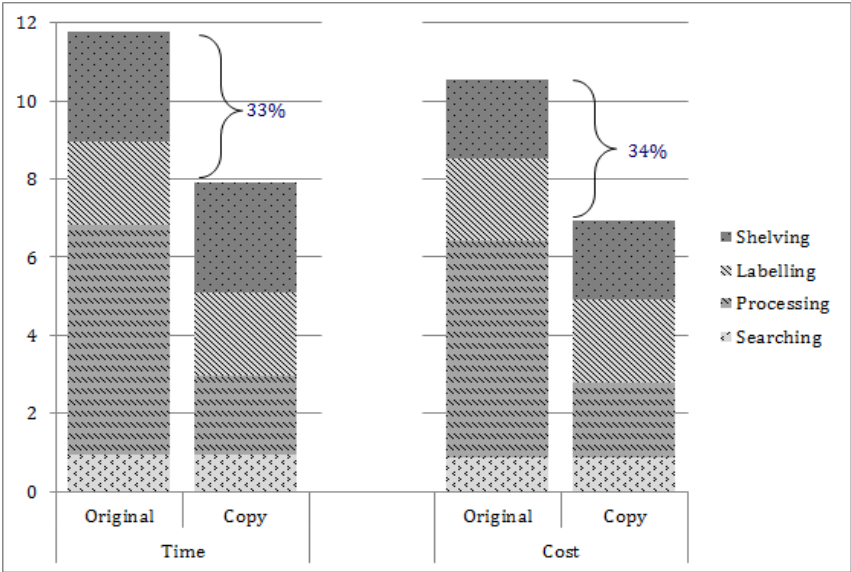


Figure B.3: Comparison of original and copy cataloging in terms of time and cost

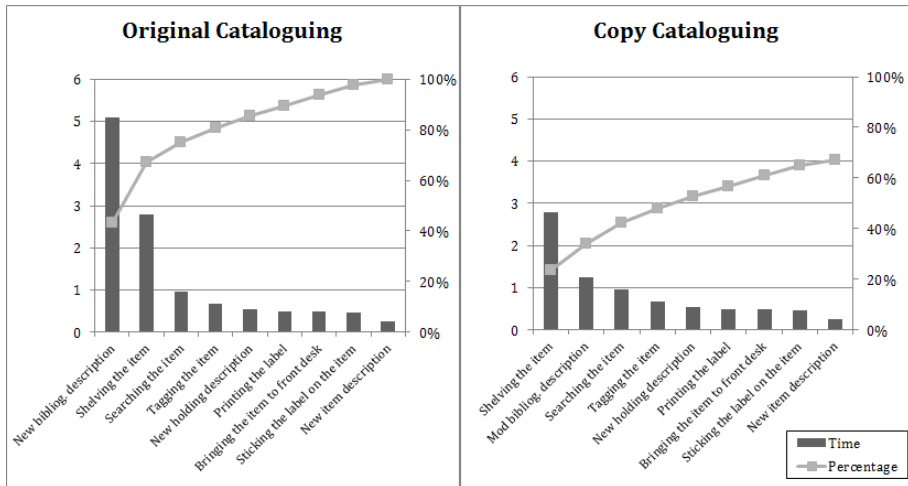


Figure B.4: Pareto chart of disaggregating costs per activity for original and copy cataloging

Moreover, time equations show that in this particular case, labelling activities such as printing and sticking labels, bringing an item to the front desk, and tagging the item, consume approximately 20-30% of the cataloging processes. Reviewing the consumed resources revealed that the cataloger is typically in charge of labelling and shelving cataloged items. A “what-if analysis” can be performed to simulate the effect of delegating these activities to student library employees (SLE). If we take as reference the SLE costs calculated by Siguenza-Guzman et al. (2014) of 0.23 €/min, then the resulting costs would be €8.63 for original cataloging, and €5.00 for copy cataloging, as shown in Table B.6. That is a cost reduction of about 18% for original cataloging and 28% for copy cataloging.

An additional improvement to the process can be to incorporate batch cataloging practices and mass retrieval rather than cataloging individual items or records (Mitchell et al., 2010). The process flow analysis enables library managers to group and improve certain activities such as searching, processing (bibliographic, holding and item descriptions), labelling and shelving. These batch activities, such as processing a bunch of items or records at one time, can be performed based on the staff availability at certain moments of the day or week. An example of these sorts of improvements on the cataloging process is shown in Table B.7. In this case, searching, labelling, and shelving are delegated to SLEs and calculated in a batch of 10 items. The duration time of searching items needs to be augmented since SLEs have no expertise in the use of LMS whereas a cataloger does have. The activity of bringing the item to the front desk is eliminated, suggesting the purchase of an extra rewriting RFID machine be located in the cataloging computer. Finally, shelving activities were recalculated based on the batching times used by Siguenza-Guzman et al. (2014) to model a returning process flow. That includes, an SLE sorting the items in the cluster (i.e., book collection divisions); and then reshelving the items. The resulting time and cost after these improvements can be seen in Table B.7.

Results of the what-if analysis indicate that by doing these changes, copy cataloging is approximately 49% less costly and 42% less time-consuming than original cataloging. In addition, for original cataloging, the obtained average time is reduced to 9.26 minutes and the cost to €7.49. This represents about 21% less-time consuming and 29% less-costly than in the real case. Most strikingly, is the case of copy cataloging in which the reduction is approximately 32% in time and 44% in cost. Therefore, the results of the what-if analysis confirm the validity of implementing these changes in the cataloging process flow as part of a best practices approach. The enhanced data flow diagram of the cataloging process is shown in Figure B.5.

Table B.6: Example of a what-if analysis applied to the original model of the cataloging process

Activity	Average Time (Minutes)	Cost per min (€/Minute)	Cost (€)	Condition	Resources
<i>Standard activities:</i>					
Searching the item on the LMS	.95	.94	.89		Cataloger + LMS + CM + GO
New holding description	.54	.94	.50		Cataloger + LMS + CM + GO
New item description	.25	.94	.24		Cataloger + LMS + CM + GO
Printing the label	.50	.55	.28		SLE + LMS + CM + GO
Sticking the label on the item	.47	.33	.15		SLE + GO
Bringing the item to front desk	.50	.33	.17		SLE + GO
Tagging the item	.67	1.02	.68		SLE + LMS + CM + GO + RM + Tag
Shelving the item	2.80	.33	.92		SLE + GO
Subtotal	6.68		3.83		
<i>Optional activities:</i>					
New bibliographic description	5.10	.94	4.80	if original_cataloging	Cataloger + LMS + CM + GO
Modify bibliographic description	1.24	.94	1.17	if copy_cataloging	Cataloger + LMS + CM + GO
Subtotal		
Total Original Cataloging	11.78		8.63		
Total Copy Cataloging	7.92		5.00		

Note. – LMS = Library Management System. SLE = Student Library Employees. GO = General Overhead. CM = Computer Maintenance. RM = RFID Maintenance.

Table B.7: Example of a what-if analysis to improve applied to the original model of the cataloging process

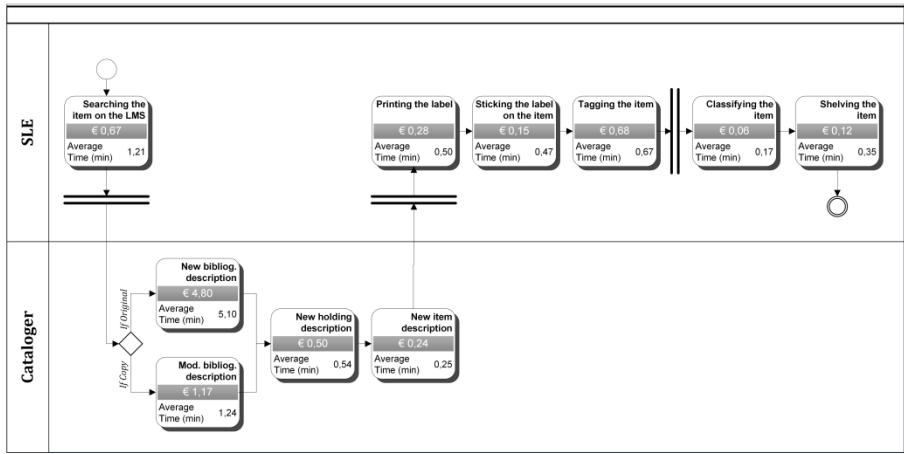
Activity	Average Time (Minutes)	Cost per min (€/Minute)	Cost (€)	Condition	Resources
<i>Searching</i>					
Searching the item on the LMS	1.21	.55	.67		SLE+ LMS + CM + GO
<i>Processing</i>					
Modify bibliographic description	1.24	.94	1.17	if copy_cataloging	Cataloger + LMS + CM + GO
New bibliographic description	5.10	.94	4.80	if original_cataloging	Cataloger + LMS + CM + GO
New holding description	.54	.94	.50		Cataloger + LMS + CM + GO
New item description	.25	.94	.24		Cataloger + LMS + CM + GO
<i>Labelling</i>					
Printing the label	.50	.55	.28		SLE + LMS + CM + GO
Sticking the label on the item	.47	.33	.15		SLE + GO
Tagging the item	.67	1.02	.68		SLE + LMS + CM + GO + RM + Tag
<i>Shelving*</i>					
Classifying the item	.17	.33	.06		SLE + GO
Shelving the item	.35	.33	.12		SLE + GO
Total Original Cataloging	9.26		7.49		
Total Copy Cataloging	5.40		3.85		

Note. – LMS = Library Management System. SLE = Student Library Employees. GO = General Overhead. CM = Computer Maintenance.

RM = RFID Maintenance.

* Shelving time taken from Siguenza-Guzman et al. (2014)

Modified Cataloging Process



Note. – LMS = Library Management System.

Figure B.5: Improvements on the cataloging process

B.5Benefits of TDABC in Cataloging Processes

The implementation of TDABC in cataloging processes at CBA demonstrated important benefits. The first important benefit shows the possibility to clearly discriminate activities regarding the time, and thus determining which activities demand more time and cost. For instance, analysis showed that labelling and shelving activities consume unnecessary resources. The second benefit is a consequence of the previous one. That is, the possibility of performing “what-if analysis” to simulate potential scenarios; for example, as labelling and shelving consumed unnecessary resources, what-if analyses were conducted. The first simulation showed that by delegating these activities to SLE staff, the library manager could reduce up to 23% of costs. The second simulation improves the previous simulation by incorporating batching processes. Results showed a reduction of approximately 37% of cataloging costs, allowing librarians to improve their processes and liberate their time to perform other specialized activities. An additional benefit is that TDABC allows benchmarking different scenarios locally and among libraries, for example, original versus copy cataloging. These results show that original cataloging in this particular scenario is 30% more time-consuming, and consequently more costly than original cataloging. Moreover, by benchmarking time and cost of activities among libraries, TDABC allows libraries to adopt policies and procedures to improve efficiency. When sharing benchmark figures, Nancy Slight-Gibney (1999) states that for other libraries, the time spent on various activities is probably more useful than the costs. In the case of benchmarking cost, the participating libraries require a common understanding of how to attribute indirect costs to their calculations and of what costs are included/excluded in order to have standard results. This problem does not occur when benchmarking the time spent on different activities. Eventually, a fourth important benefit is that TDABC allows justifying decisions and choices. TDABC allows both managers and staff to better understand alternative options and accept the need for change; for example, the decision of transferring responsibilities from catalogers to SLE staff, as well as, structuring batch activities (processing, labelling and shelving). In fact, library managers should constantly analyze their cost information and keep their models updated in order to redesign workflows efficiently and effectively, as well as to reallocate resources and tasks.

Nevertheless, Siguenza-Guzman et al. (2013; 2014) suggest important considerations to be borne in mind at implementing TDABC in academic libraries, namely: 1) the resource intensity of data collection to gather the time duration of activities, as well as to document the activity flows; 2) the need of a dedicated software tool to keep the flows updated and consequently to facilitate long-term maintenance; and 3) the commitment of library managers and staff during the data collection.

B.6Conclusions

In this paper, the TDABC implementation was described in two main cataloging activities, namely, original and copy cataloging. These processes were selected because they are considered to be a part of the core activities of a library to manage its collection, but are also resource intensive. Based on our findings, we can conclude that TDABC is a quicker and easier way of calculating cataloging savings. In fact, TDABC is a useful method to perform cost analysis in cataloging processes, and consequently provides valuable data for managerial decisions. The TDABC implementation provided library managers with important information about cataloging costs and performance measurements; and guided decisions concerning resource allocation and process improvements. For example, based on the obtained results, the library manager decided to delegate certain activities and define a set of batch activities. This case study, therefore, shows significant contributions to the literature on the implementation of advanced cost models for library processes, and more precisely for cataloging activities.

A potential direction for future research is to expand this study to different cataloging activities such as cooperative, contract, and outsourcing cataloging. TDABC can also offer the possibility to discuss how these trends in cataloging processes affect cataloging units. TDABC will allow to examine whether these trends really provide an opportunity for catalogers to spend their time on other cataloging activities, such as enhancing existing records, or not. Another interesting area for future research would be a similar analysis for cataloging audio-visual items and other special materials such as old books, non-print materials, and maps. Utilizing TDABC to benchmark libraries for “best practices” is an additional prospect for future analysis.

References

- ACRL Research Planning and Review Committee. (2010). 2010 top ten trends in academic libraries A review of the current literature. *College & Research Libraries News*, 71(6), 286–292.
- Beall, J., & Kafadar, K. (2004). The effectiveness of copy cataloging at eliminating typographical errors in shared bibliographic records. *Library Resources & Technical Services*, 48(2), 92–101.
- Campus Bibliotheek Arenberg. (2013, December 23). CBA - Campusbibliotheek Arenberg. Retrieved January 22, 2014, from <https://bib.kuleuven.be/cba/cba-english/home>
- Charbonneau, M. D. (2005). Production benchmarks for catalogers in academic libraries. *Library Resources & Technical Services*, 49(1), 40–48.
- Ching, S. H., Leung, M. W., Fidow, M., & Huang, K. L. (2008). Allocating costs in the business operation of library consortium: The case study of Super e-Book Consortium. *Library Collections, Acquisitions, and Technical Services*, 32(2), 97–103. <http://doi.org/16/j.lcats.2008.08.013>
- Cooper, R., & Kaplan, R. S. (1988). Measure costs right: Make the right decision. *Harvard Business Review*, 66(5), 96–103.
- Cottrell, T. (Terry). (2012). Three phantom budget cuts and how to avoid them. *Bottom Line: Managing Library Finances*, The, 25(1), 16–20. <http://doi.org/10.1108/08880451211229171>
- Ellis-Newman, J. (2003). Activity-Based Costing in user services of an academic library. *Library Trends*, 51(3), 333–348.
- Ellis-Newman, J., Izan, H., & Robinson, P. (1996). Costing support services in universities: An application of activity-based costing. *Journal of Institutional Research in Australasia*, 5(1), 75–86.

- Ellis-Newman, J., & Robinson, P. (1998). The cost of library services: Activity-based costing in an Australian academic library. *The Journal of Academic Librarianship*, 24(5), 373–379. [http://doi.org/16/S0099-1333\(98\)90074-X](http://doi.org/16/S0099-1333(98)90074-X)
- Everaert, P., Bruggeman, W., Sarens, G., Anderson, S. R., & Levant, Y. (2008). Cost modeling in logistics using time-driven ABC: Experiences from a wholesaler. *International Journal of Physical Distribution & Logistics Management*, 38(3), 172–191. <http://doi.org/10.1108/09600030810866977>
- Goddard, A., & Ooi, K. (1998). Activity-Based Costing and central overhead cost allocation in universities: A case study. *Public Money and Management*, 18(3), 31–38. <http://doi.org/10.1111/1467-9302.00124>
- Hoozée, S., Vermeire, L., & Bruggeman, W. (2012). The impact of refinement on the accuracy of Time-Driven ABC. *Abacus*, 48(4), 439–472. <http://doi.org/10.1111/j.1467-6281.2012.00366.x>
- Kaplan, R. S., & Anderson, S. R. (2004). Time-Driven Activity-Based Costing - Tool Kit. *Harvard Business Review*, (82), 131–138.
- Kaplan, R. S., & Anderson, S. R. (2007). Time-Driven Activity-Based Costing: A simpler and more powerful path to higher profits. Boston, MA, USA: Harvard Business School Press.
- Kaplan, R. S., & Cooper, R. (1998). Cost & Effect: Using integrated cost systems to drive profitability and performance. Boston, MA, USA: Harvard Business School Press.
- Kont, K.-R. (2011). New cost accounting models in measuring of library employees' performance. *Library Management*, 33(1/2), 50–65. <http://doi.org/10.1108/01435121211203310>
- Manaf, Z. A., & Rahman, R. A. (2006). Examining the quality of National Library of Malaysia (NLM) cataloguing in publication (CIP) records. *Library Review*, 55(6), 363–373. <http://doi.org/10.1108/00242530610674776>
- McCain, C., & Shorten, J. (2002). Cataloging efficiency and effectiveness. *Library Resources & Technical Services*, 46(1), 23–31. <http://doi.org/10.5860/lrts.46n1.23>
- Mitchell, A. M., Thompson, J. M., & Wu, A. (2010). Agile cataloging: Staffing and skills for a bibliographic future. *Cataloging & Classification Quarterly*, 48(6-7), 506–524. <http://doi.org/10.1080/01639374.2010.496308>
- Morris, D. E. (1992). Staff time and costs for cataloging. *Library Resources & Technical Services*, 36(1), 79–95.
- Morris, D. E., Hobert, C. B., Osmus, L., & Wool, G. (2000). Cataloging staff costs revisited. *Library Resources & Technical Services*, 44(2), 70–83.
- Morris, D. E., & Wool, G. (1999). Cataloging: Librarianship's best bargain. *Library Journal*, 124(11), 44–46.
- Novak, D. D., Paulos, A., & Clair, G. S. (2011). Data-driven budget reductions: A case study. *The Bottom Line: Managing Library Finances*, 24(1), 24–34. <http://doi.org/10.1108/08880451111142015>
- Pernot, E., Roodhooft, F., & Van den Abbeele, A. (2007). Time-Driven Activity-Based Costing for inter-library services: A case study in a university. *The Journal of Academic Librarianship*, 33(5), 551–560. <http://doi.org/16/j.acalib.2007.06.001>
- Reitz, J. M. (2004). ABC-CLIO - ODLIS. ODLIS - Online dictionary for library and information science (p. 788). *Libraries Unlimited*. Retrieved from http://www.abc-clio.com/odlis/odlis_a.aspx
- Roberts, S. A. (2003). Financial management of libraries: Past trends and future prospects. *Library Trends*, 51(3), 462–493.
- Siguenza-Guzman, L., Holans, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2013). Towards a holistic analysis tool to support decision-making in libraries. In *Proceedings of the IATUL Conferences* (p. Paper 29). Cape Town, South Africa: Purdue e-Pubs. Retrieved from <http://docs.lib.purdue.edu/iatul/2013/papers/29/>
- Siguenza-Guzman, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2013). Recent evolutions in costing systems: A literature review of Time-Driven Activity-Based Costing. *ReBEL - Review of Business and Economic Literature*, 58(1), 34–64.
- Siguenza-Guzman, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2014). Using Time-Driven Activity-Based Costing to support library management decisions: A case study for lending and returning processes. *Library Quarterly: Information, Community, Policy*, 84(1), 76–98. <http://doi.org/10.1086/674032>

- Slight-Gibney, N. (1999). How far have we come? Benchmarking time and costs for monograph purchasing. *Library Collections, Acquisitions, and Technical Services*, 23(1), 47-59. [http://doi.org/10.1016/S1464-9055\(98\)00125-0](http://doi.org/10.1016/S1464-9055(98)00125-0)
- Stouthuysen, K., Swiggers, M., Reheul, A.-M., & Roodhooft, F. (2010). Time-Driven Activity-Based Costing for a library acquisition process: A case study in a Belgian University. *Library Collections, Acquisitions, and Technical Services*, 34(2-3), 83-91. <http://doi.org/10.1016/j.lcats.2010.05.003>
- Vazakidis, A., & Karagiannis, I. (2009). Activity-Based Management and traditional costing in tourist enterprises (A hotel implementation model). *Operational Research*, 11(2), 123-147. <http://doi.org/10.1007/s12351-009-0049-3>
- Wegmann, G., & Nozile, S. (2009). The Activity-Based Costing method developments: State-of-the art and case study. *The IUP Journal of Accounting Research and Audit Practices*, 8(1), 7-22. <http://doi.org/oai:halshs.archives-ouvertes.fr:halshs-00482137>
- Yilmaz, R. (2008). Creating the profit focused organization using Time-Driven Activity Based Costing. In *EABR & TLC Conferences Proceedings* (p. 8). Salzburg, Austria: Clute Institute for Academic Research. Retrieved from <http://rifatyilmaz.net/SLZ08-191.pdf>

Appendix C: TD-ABC-D: Time-Driven Activity-Based Costing Software for libraries

Siguenza-Guzman, L., Cabrera, P., & Cattrysse, D. (2014, August). TD-ABC-D: Time-Driven Activity-Based Costing Software for Libraries. Poster presented at the 80th IFLA General Conference and Assembly, Lyon, France.

C.1 Introduction

Time-Driven Activity-Based Costing (TDABC) is a relatively new costing method, which is gradually gaining acceptance in libraries, thanks to its simplicity and rapid implementation. TDABC only requires the unit cost of supplying resource capacity and the time estimation to perform activities (Kaplan & Anderson, 2007); allowing library managers with no experience in accounting to quickly conduct cost studies. Up to now, four important studies on TDABC in libraries have been applied to very specific processes such as inter-library loan (Pernot, Roodhooft, & Van den Abbeele, 2007), acquisition (Stouthuysen, Swiggers, Reheul, & Roodhooft, 2010), circulation (Siguenza-Guzman, Van den Abbeele, Vandewalle, Verhaaren, & Cattrysse, 2014), and cataloging (Siguenza-Guzman, Van Den Abbeele, & Cattrysse, 2014). In these case studies, three specific TDABC advantages in libraries are highlighted: disaggregate values per activity, compare different scenarios, and justify decisions and choices (Siguenza-Guzman et al., 2013).

The aim of this project is to provide a web-based software tool TD-ABC-D for TDABC analysis in libraries processes. Its development has been promoted and coordinated under a PhD program of the University of Leuven (Belgium) with the support of the University of Cuenca (Ecuador), VLIR-UOS and SENESCYT. TD-ABC-D has been implemented as an additional module of the integrated library management software called ABCD. (in Spanish: 'Automatización de Bibliotecas y Centros de Documentación').


The main features of TD-ABC-D²¹ are its simple user interface, easy drag & drop functionality to set up process flows, dynamic and user friendliness, as well as, highly accuracy to provide results. TD-ABC-D, that stands for the TDABC module in ABCD systems, has been developed under the philosophy of "Free and Open Source Software (FOSS)", using freely available technologies like PHP

²¹ <http://tdabc-libraries.com/>

as server-side scripting language, Apache as web server, MySQL as database, and several tools for web development. The aim of this poster is to describe the three main TD-ABC-D modules: TDABC, process simulation and benchmarking; to present its main technical characteristics and planed future developments.

C.2 Poster

KU LEUVEN



TD-ABC-D: Time-Driven Activity-Based Costing Software for Libraries

Lorena Siguenza-Guzman^{1,2}, Paul Cabrera Encalada³, Dirk Cattrysse¹

(1) Centre for Industrial Management, Traffic & Infrastructure, KU Leuven (BELGIUM)

(2) Department of Computer Science, University of Cuenca (ECUADOR)

(3) Faculty of Engineering, University of Cuenca (ECUADOR)

SETTING

- Periodic shortages and limited budgets obligate libraries to provide high-quality services at lower costs.
- More than ever, library managers must keep their activities, resources and costs under control.
- TDABC is a fast and easy way to accurately calculate the cost of a wide range of processes and services.
- TDABC only requires the cost of resource consumption and the time consumed by the activities.


OBJECTIVES

- To provide a dedicated software tool for costing analysis in library processes.
- To develop TD-ABC-D as an additional Web-based module of ABCD (Integrated Library Automation Software).
- To minimize development costs by using free and open-source software (FOSS).

MODULES

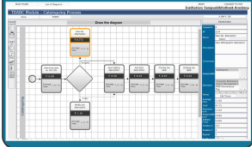
1

TDABC module




2

Risk Analysis



3


Benchmarking





RESULTS


- Easy and simple interface to maintain processes, responsible and resources.
- Interactive process cost information.
- Easy drag & drop functionality to create process flows.
- Possibility to evaluate different scenarios and risks by performing what-if-analysis.
- Benchmark of process and service cost and time among institutions and branches.

INSTITUTIONS









www.tdabc-libraries.com

info@tdabc-libraries.com

238

References

- Kaplan, R. S., & Anderson, S. R. (2007). *Time-Driven Activity-Based Costing: A simpler and more powerful path to higher profits*. Boston, MA, USA: Harvard Business School Press.
- Pernot, E., Roodhooft, F., & Van den Abbeele, A. (2007). Time-Driven Activity-Based Costing for inter-library services: A case study in a university. *The Journal of Academic Librarianship*, 33(5), 551–560. <http://doi.org/16/j.acalib.2007.06.001>
- Siguenza-Guzman, L., Holans, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2013). Towards a holistic analysis tool to support decision-making in libraries. In *Proceedings of the IATUL Conferences* (p. Paper 29). Cape Town, South Africa: Purdue e-Pubs. Retrieved from <http://docs.lib.purdue.edu/iatul/2013/papers/29/>
- Siguenza-Guzman, L., Van Den Abbeele, A., & Cattrysse, D. (2014). Time-Driven Activity-Based Costing Systems for Cataloguing Processes: A Case Study. *LIBER Quarterly*, 23(3), 160–186.
- Siguenza-Guzman, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2014). Using Time-Driven Activity-Based Costing to support library management decisions: A case study for lending and returning processes. *Library Quarterly: Information, Community, Policy*, 84(1), 76–98. <http://doi.org/10.1086/674032>
- Stouthuysen, K., Swiggers, M., Reheul, A.-M., & Roodhooft, F. (2010). Time-Driven Activity-Based Costing for a library acquisition process: A case study in a Belgian University. *Library Collections, Acquisitions, and Technical Services*, 34(2-3), 83–91. <http://doi.org/16/j.lcats.2010.05.003>

Appendix D: **Using Time-Driven Activity-**

Based Costing to identify best practices in

academic libraries: Costing tables

Siguenza-Guzman, L., Auquilla, A., Van Den Abbeele, A., & Cattrysse, D. (2015). Using Time-Driven Activity-Based Costing to identify best practices in academic libraries: Costing Tables ([Secondary document]) (p. 10p.). KU Leuven. Retrieved from <https://lirias.kuleuven.be/handle/123456789/495588>

D.1Introduction

The present document contains detailed information on the costing tables discussed in the article “Using Time-Driven Activity-Based Costing to identify best practices in academic libraries” (Chapter 5). The total costs and time incurred in nine different processes by libraries 1 and 2 are provided in tables C.1–C.18. The processes analyzed are the following: acquisition of journals, original and copy cataloging, lending and returning items, requesting closed stack items, and interlibrary loan processes that are outgoing and incoming request of digital and printed items. Although, all costing tables are detailed in this Appendix, the corresponding analysis is provided in the discussion section of the Chapter 5.

Each cost table contains seven columns. The first column lists the standard activities identified in the process, while the second column indicates all the particular activities performed in the process by each library. The third column shows the average time per activity and the fourth column contains the cumulative time per standard activity. The fifth column indicates the accumulated costs of each resource group, the sixth column calculates the resulting cost incurred in the activity, and the last column shows the resulting cost per standard activity.

APPENDIX D

Table D.1: Journals acquisition process cost table (Library 1)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Selecting	Selecting journals	0.95	1.21	0.86 ^a	0.81	1.04
	Check on the system	0.26		0.85 ^b	0.22	
Requesting the journal	Put the order on the request	1.06	1.28	0.85	0.90	1.07
	Send email with the request	0.22		0.75 ^c	0.17	
Ordering	Order with the bib description	0.39	0.39	0.85	0.33	0.33
Purchasing	Register the invoice in LMS	1.77	4.76	0.85	1.50	3.81
	Pick up the invoice report	1.43		0.85	1.21	
	Put invoice in delivery room	1.56		0.70 ^d	1.10	
Receiving materials	Check the journals	0.21	5.37	0.70	0.15	3.89
	Register the journal in LMS	0.64		0.85	0.54	
	Label journal	0.12		0.70	0.08	
	Shelve the journal temporally	0.04		0.70	0.03	
	Put journal in the delivery room	1.56		0.70	1.10	
	Shelve the item	2.80		0.71 ^e	1.99	
<i>Total</i>			13.01			10.13

^a 0.86 = *Cataloger + LMS + Computers*
^b 0.85 = *Acquirer + LMS + Computers*
^c 0.75 = *Acquirer + Computers*
^d 0.70 = *Acquirer*
^e 0.71 = *Cataloger*

Table D.2: Journals acquisition process cost table (Library 2)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Selecting	Select journals	1.00	1.00	1.79 ^a	1.79	1.79
Requesting the journal	Log in to LMS	0.55	3.38	0.85 ^b	0.47	2.87
	Connect to Financial System	1.00		0.85	0.85	
	Connect to journal deliver	0.42		0.85	0.35	
	Consult the catalog	1.42		0.85	1.20	
Ordering	Add the item to shopping basket	0.08	2.58	0.85	0.07	2.19
	Check if order ok	1.75		0.85	1.49	
	Order & pay	0.75		0.85	0.64	
Purchasing	Connect to Financial System	1.00	1.50	0.85	0.85	1.23
	Read e-mail	0.50		0.75 ^c	0.38	
Receiving materials	Register the journal in LMS	5.00	12.00	0.86 ^d	4.29	10.50
	Label journal	3.33		1.08 ^e	3.61	
	Shelve the item	3.67		0.71 ^f	2.60	
<i>Total</i>			20.47			18.59

^a 1.79 = *Director + Computers*
^b 0.85 = *Acquirer + LMS + Computers*
^c 0.75 = *Acquirer + Computers*
^d 0.86 = *Cataloger + LMS + Computers*
^e 1.08 = *Cataloger + RFID + RFID_{maint} + LMS + Computers*
^f 0.71 = *Cataloger*

Table D.3: Original cataloging process cost table (Library 1)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Searching	Search the item on the LMS	0.95	0.95	0.86 ^a	0.81	0.81
Processing	New bibliographic description	4.16	4.91	0.86	3.57	4.21
	New holding description	0.51		0.86	0.44	
	New item description	0.24		0.86	0.21	
Labeling	Print label	0.50	2.15	0.86	0.43	1.95
	Stick label on the item	0.48		0.71 ^b	0.34	
	Bring item to front desk	0.50		0.71	0.36	
	Tag the item	0.67		1.23 ^c	0.82	
Shelving	Shelve the item	2.80	2.80	0.71	1.99	1.99
<i>Total</i>			10.81			8.96

^a 0.86 = *Cataloger* + *LMS* + *Computers*
^b 0.71 = *Cataloger*
^c 1.23 = *Cataloger* + *RFID* + *RFID_{maint}* + *LMS* + *Computers*

Table D.4: Original cataloging process cost table (Library 2)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Searching	Log in to LMS	0.90	1.27	0.86 ^a	0.77	1.09
	Search database	0.37		0.86	0.31	
Processing	Create a new record	0.38	2.71	0.86	0.33	2.33
	New bibliographic description	0.90		0.86	0.77	
	Save the register	0.15		0.86	0.13	
	New holding description	1.28		0.86	1.10	
Labeling	Print barcode number	0.17	1.88	0.86	0.14	1.55
	Bring item to front desk	0.42		0.71 ^b	0.30	
	Transfer barcode to RFID	0.27		1.23 ^c	0.33	
	Tag the item	0.20		0.71	0.14	
	Print label	0.33		0.86	0.29	
	Stick label on the item	0.50		0.71	0.36	
Shelving	Shelve the item	5.17	5.17	0.42 ^d	2.15	2.15
<i>Total</i>			11.03			7.11

^a 0.86 = *Cataloger* + *LMS* + *Computers*
^b 0.71 = *Cataloger*
^c 1.23 = *Cataloger* + *RFID* + *RFID_{maint}* + *LMS* + *Computers*
^d 0.42 = *SLE*

Table D.5: Copy cataloging process cost table (Library 1)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Searching	Search the item on the LMS	0.95	0.95	0.86 ^a	0.81	0.81
Processing	Modify bibliog. description	1.24	1.99	0.86	1.06	1.71
	New holding description	0.51		0.86	0.44	
	New item description	0.24		0.86	0.21	
Labeling	Print label	0.50	2.15	0.86	0.43	1.95
	Stick label on the item	0.48		0.71 ^b	0.34	
	Bring item to front desk	0.50		0.71	0.36	
	Tag the item	0.67		1.23 ^c	0.82	
Shelving	Shelve the item	2.80	2.80	0.71	1.99	1.99
<i>Total</i>			7.89			6.46

^a 0.86 = Cataloger + LMS + Computers^b 0.71 = Cataloger^c 1.23 = Cataloger + RFID + RFID_{maint} + LMS + Computers**Table D.6:** Copy cataloging process cost table (Library 2)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Searching	Log in to LMS	0.90	1.27	0.86 ^a	0.77	1.09
	Search database	0.37		0.86	0.31	
Processing	Validate the information	0.63	1.92	0.86	0.54	1.64
	New holding description	1.28		0.86	1.10	
Labeling	Print barcode number	0.17	1.88	0.86	0.14	1.55
	Bring item to front desk	0.42		0.71 ^b	0.30	
	Transfer barcode to RFID	0.27		1.23 ^c	0.33	
	Tag the item	0.20		0.71	0.14	
	Print label	0.33		0.86	0.29	
	Stick label on the item	0.50		0.71	0.36	
Shelving	Shelve the item	5.17	5.17	0.42	2.15	2.15
<i>Total</i>			10.23			6.43

^a 0.86 = Cataloger + LMS + Computers^b 0.71 = Cataloger^c 1.23 = Cataloger + RFID + RFID_{maint} + LMS + Computers^d 0.42 = SLE**Table D.7:** Lending process cost table (Library 1)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Searching	Consult the catalog	1.36	1.36	0.15 ^a	0.20	0.20
Borrowing	Use self-check equipment	0.38	0.49	0.07 ^b	0.03	0.04
	Print the receipt	0.11		0.07	0.01	
<i>Total</i>			1.85			0.24

^a 0.15 = LMS + Computers^b 0.07 = RFID_{maint}

Table D.8: Lending process cost table (Library 2)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Searching	Consult the catalog	0.75	0.75	0.15 ^a	0.11	0.11
Borrowing	Use self-check equipment	0.35	0.43	0.07 ^b	0.03	0.03
	Print the receipt	0.08		0.07	0.01	
<i>Total</i>			1.18			0.14

^a 0.15 = *LMS* + *Computers*
^b 0.07 = *RFID_{maint}*

Table D.9: Returning process cost table (Library 1)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Returning	Return item to self-check equipment	0.38	0.48	0.07 ^a	0.03	0.04
	Print the receipt	0.11		0.07	0.01	
Classifying	Go to the robot	0.03	0.28	0.42 ^b	0.01	0.12
	Return item to returning robot	0.08		0.49 ^c	0.04	
	Classify the item	0.17		0.42	0.07	
Shelving	Shelve the item	0.35	0.35	0.42	0.15	0.15
<i>Total</i>			1.11			0.31

^a 0.07 = *RFID_{maint}*
^b 0.42 = *SLE*
^c 0.49 = *SLE* + *RFID_{maint}*

Table D.10: Returning process cost table (Library 2)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Returning	Return item to self-check equipment	0.23	0.41	0.07 ^a	0.03	0.09
	Print the receipt	0.08		0.07	0.01	
	Return item to book truck	0.10		0.67 ^b	0.07	
Classifying	Classify the item	0.50	0.50	0.42 ^c	0.21	0.21
Shelving	Shelve the item	1.03	1.03	0.42	0.43	0.43
<i>Total</i>			1.94			0.74

^a 0.07 = *RFID_{maint}*
^b 0.67 = *Circulation*
^c 0.42 = *SLE*

Table D.11: Requesting closed stack items process (Library 1)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Searching	Consult the catalog	1.36	1.36	0.15 ^a	0.20	0.20
Requesting	Fill request form (online)	0.68	2.68	0.15	0.10	1.72
	Print the request	2.00		0.81 ^b	1.62	
Retrieving	Send the print demand	0.15	1.60	0.67 ^c	0.10	0.89
	Get the print demand	0.15		0.53 ^d	0.08	
	Get the item from stack	1.00		0.53	0.53	
	Send item to main desk through lift	0.15		0.53	0.08	
	Get the item from the lift	0.15		0.67	0.10	
Delivering	Get the item from circulation desk	0.17	0.17	0.67	0.11	0.11
Shelving	Delete the request	0.32	1.20	0.81	0.26	0.71
	Return the item to the lift	0.20		0.67	0.13	
	Get the item from the lift	0.15		0.53	0.08	
	Pre-shelve the item	0.18		0.53	0.10	
	Shelve the item	0.35		0.42 ^e	0.15	
Total			7.01			3.64

^a 0.15 = LMS + Computers
^b 0.81 = Circulation + LMS + Computers
^c 0.67 = Circulation
^d 0.53 = Closed_stack_resp
^e 0.42 = SLE

Table D.12: Requesting closed stack items process (Library 2)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Searching	Consult the catalog	0.75	0.75	0.15 ^a	0.11	0.11
Requesting	Fill request form (by hand)	1.27	1.27	0.10 ^b	0.13	0.13
Retrieving	Get the item from the stack	5.07	7.74	0.67 ^c	3.37	5.15
	Bring to the library	2.67		0.67	1.78	
Delivering	Get the item from circulation desk	0.17	0.17	0.67	0.11	0.11
Shelving	Return item to archive	4.35	4.35	0.67	2.89	2.89
Total			14.28			8.39

^a 0.15 = LMS + Computers
^b 0.10 = General overhead
^c 0.67 = Circulation

Table D.13: ILL outgoing request process (Library 1)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Ordering	Send an ILL request	0.20	6.65	0.10 ^a	0.02	5.20
	Check request	0.42		0.67 ^b	0.28	
	Look for supplier	2.42		0.81 ^c	1.97	
	Check username on LMS	0.86		0.81	0.70	
	Order it	2.75		0.81	2.23	
Receiving	Check the LMS	0.76	1.15	0.81	0.62	0.93
	Receive the pdf file	0.39		0.81	0.32	
Delivering	Enter into the database	0.45	2.87	0.81	0.37	2.22
	Send the pdf to the customer	0.48		0.72 ^d	0.34	
	Check the ILL system	0.53		0.81	0.43	
	Fill in Excel file	0.67		0.72	0.48	
	Close the request	0.74		0.81	0.60	
<i>Total</i>			10.67			8.35

^a 0.10 = *LMS*
^b 0.67 = *ILL_resp*
^c 0.81 = *ILL_resp* + *LMS* + *Computers*
^d 0.72 = *ILL_resp* + *Computers*

Table D.14: ILL outgoing request process (Library 2)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Ordering	Check University network	1.33	6.25	0.81 ^a	1.08	5.08
	Check Belgian ILL network	1.75		0.81	1.42	
	Check German ILL network	1.50		0.81	1.22	
	Check British Library network	1.67		0.81	1.36	
Receiving	Receive the pdf file	0.58	0.91	0.81	0.47	0.71
	Rename the pdf file	0.33		0.72 ^b	0.24	
Delivering	Notify requester by email	0.83	1.99	0.72	0.59	1.46
	Remove item from waiting list	0.33		0.81	0.27	
	Send the pdf to the customer	0.83		0.72	0.59	
<i>Total</i>			9.15			7.24

^a 0.81 = *ILL_resp* + *LMS* + *Computers*
^b 0.72 = *ILL_resp* + *Computers*

APPENDIX D

Table D.15: ILL incoming request process for digital items (Library 1)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Requesting	Print the request	0.17	0.17	0.81 ^a	0.14	0.14
Retrieving	Looking up the request - online	1.80	1.80	0.81	1.46	1.46
Delivering	Send the pdf to the customer	1.37	1.37	0.72 ^b	0.98	0.98
Pricing	Fill the price and close the request	0.52	0.71	0.81	0.42	0.55
	Sort the request	0.14		0.67 ^c	0.09	
	Make a copy of the request	0.05		0.72	0.04	
<i>Total</i>			4.05			3.13

^a 0.81 = *ILL_resp* + *LMS* + *Computers*

^b 0.72 = *ILL_resp* + *Computers*

^c 0.67 = *ILL_resp*

Table D.16: ILL incoming request process for digital items (Library 2)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Requesting	Check Impala system	1.50	1.50	0.81 ^a	1.22	1.22
Retrieving	Looking up the request - online	0.50	0.92	0.81	0.41	0.75
	Fill in supplier's details	0.42		0.81	0.34	
Delivering	Send the pdf to the server	0.67	0.67	0.81	0.54	0.54
Pricing	Fill the price and close the request	0.83	0.83	0.81	0.67	0.67
<i>Total</i>			3.92			3.18

^a 0.81 = *ILL_resp* + *LMS* + *Computers*

Table D.17: ILL incoming request process for printed items (Library 1)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Requesting	Print the request	0.17	0.17	0.81 ^a	0.14	0.14
Retrieving	Looking up the request - printed	1.13	2.62	0.81	0.92	1.54
	Search the item	1.49		0.42 ^b	0.62	
Delivering	Scanning the item	3.71	7.84	0.47 ^c	1.73	4.00
	Check the pdf	2.76		0.47	1.29	
	Send the pdf to the customer	1.37		0.72 ^d	0.98	
Pricing	Fill the price and close the request	0.52	0.71	0.81	0.42	0.55
	Sort the request	0.14		0.67 ^e	0.09	
	Make a copy	0.05		0.72	0.04	
<i>Total</i>			11.34			6.23

^a 0.81 = *ILL_resp* + *LMS* + *Computers*

^b 0.42 = *SLE*

^c 0.47 = *SLE* + *Computers*

^d 0.72 = *ILL_resp* + *Computers*

^e 0.67 = *ILL_resp*

Table D.18: ILL incoming request process for printed items (Library 2)

Standard Activity	Activity	Average Time (min)	Time per standard activity (min)	Cost (€/min)	Total Cost per activity(€)	Cost per standard activity(€)
Requesting	Check ILL system	1.50	2.58	0.81 ^a	1.22	1.95
	Sort the request	0.08		0.67 ^b	0.06	
	Make a copy	0.33		0.72 ^c	0.24	
	Bring the request to the front desk	0.67		0.67	0.44	
Retrieving	Looking up the request - printed	0.33	2.33	0.67	0.22	1.05
	Search the item	2.00		0.42 ^d	0.83	
Delivering	Send the pdf to the server	0.67	3.67	0.81	0.54	1.99
	Sort the items	0.75		0.42	0.31	
	Scan the item	1.50		0.47 ^e	0.70	
	Put the item on server	0.33		0.42	0.14	
	Check the pdf file	0.25		0.72	0.18	
	Rename the pdf file	0.17		0.72	0.12	
Pricing	Fill the price and close the request	0.83	0.83	0.81	0.67	0.67
<i>Total</i>			9.41			5.67

^a 0.81 = $ILL_{resp} + LMS + Computers$
^b 0.67 = ILL_{resp}
^c 0.72 = $ILL_{resp} + Computers$
^d 0.42 = SLE
^e 0.47 = $SLE + Computers$

Appendix E: **References of the empirical studies implementing data mining techniques in libraries**

Siguenza-Guzman, L., Saquicela, V., Avila-Ordóñez, E., Vandewalle, J., & Cattrysse, D. (2015). Literature review of data mining applications in academic libraries. The Journal of Academic Librarianship, <http://dx.doi.org/10.1016/j.acalib.2015.06.007>.

E.1Introduction

The present document contains detailed list of the selected articles utilized for the article "Literature Review of Data Mining Applications in Academic Libraries" (Chapter 7).

- Ahmad, P., Brogan, M., & Johnstone, M. N. (2014). The e-book power user in academic and research libraries: Deep log analysis and user customisation. *Australian Academic & Research Libraries*, 45(1), 35–47. <http://doi.org/10.1080/00048623.2014.885374>
- Blecic, D. D., Bangalore, N. S., Dorsch, J. L., Henderson, C. L., Koenig, M. H., & Weller, A. C. (1998). Using transaction log analysis to improve OPAC retrieval results. *College & Research Libraries*, 59(1), 39–50. <http://doi.org/10.5860/crl.59.1.39>
- Bollen, J., & Luce, R. (2002). Evaluation of digital library impact and user communities by analysis of usage patterns. *D-Lib Magazine*, 8(6). <http://doi.org/10.1045/june2002-bollen>
- Bracke, P. J. (2004). Web usage mining at an academic health sciences library: an exploratory study. *Journal of the Medical Library Association*, 92(4), 421–428.
- Decker, R., & Hermelbracht, A. (2006). Planning and evaluation of new academic library services by means of Web-based conjoint analysis. *The Journal of Academic Librarianship*, 32(6), 558–572. <http://doi.org/10.1016/j.acalib.2006.06.016>
- Decker, R., & Höppner, M. (2006). Strategic planning and customer intelligence in academic libraries. *Library Hi Tech*, 24(4), 504–514. <http://doi.org/10.1108/07378830610715374>
- Emmons, M., & Wilkinson, F. C. (2011). The Academic Library Impact on Student Persistence. *College & Research Libraries*, crl-74r1.
- Fagan, J. C. (2014). The Effects of Reference, Instruction, Database Searches, and Ongoing Expenditures on Full-text Article Requests: An Exploratory Analysis. *The Journal of Academic Librarianship*, 40(3–4), 264–274. <http://doi.org/10.1016/j.acalib.2014.04.002>
- Finnell, J., & Fontane, W. (2010). Reference Question Data Mining: A Systematic Approach to Library Outreach. *Reference & User Services Quarterly*, 49(3), 278–286.

- Hájek, P., & Stejskal, J. (2014). Library user behavior analysis - Use in economics and management. *Wseas Transactions on Business and Economics*, 11, 107–116.
- Kao, S.-C., Chang, H.-C., & Lin, C.-H. (2003). Decision support for the academic library acquisition budget allocation via circulation database mining. *Information Processing & Management*, 39(1), 133–147. [http://doi.org/10.1016/S0306-4573\(02\)00019-5](http://doi.org/10.1016/S0306-4573(02)00019-5)
- King, J. D., Li, Y., Tao, X., & Nayak, R. (2007). Mining world knowledge for analysis of search engine content. *Web Intelligence and Agent Systems*, 5(3), 233–253.
- Koulouris, A., & Kapidakis, S. (2012). Policy route map for academic libraries' digital content. *Journal of Librarianship and Information Science*, 44(3), 163–173. <http://doi.org/10.1177/0961000612444299>
- Leonard, M. F., Haas, S. C., & Kisling, V. N. (2010). Metrics and science monograph collections at the marston science library, University of Florida. *Issues in Science and Technology Librarianship*, 62. Retrieved from <http://dialnet.unirioja.es/servlet/articulo?codigo=3706724>
- Li, X. (2014). An Algorithm for Mining Frequent Itemsets from Library Big Data. *Journal of Software*, 9(9), 2361–2365. <http://doi.org/10.4304/jsw.9.9.2361-2365>
- Nicholas, D., Huntington, P., Monopoli, M., & Watkinson, A. (2006). Engaging with scholarly digital libraries (publisher platforms): The extent to which “added-value” functions are used. *Information Processing & Management*, 42(3), 826–842. <http://doi.org/10.1016/j.ipm.2005.03.019>
- Nicholson, S. (2003). Bibliomining for automated collection development in a digital library setting: Using data mining to discover Web-based scholarly research works. *Journal of the American Society for Information Science and Technology*, 54(12), 1081–1090. <http://doi.org/10.1002/asi.10313>
- Papatheodorou, C., Kapidakis, S., Sfakakis, M., & Vassiliou, A. (2003). Mining user communities in digital libraries. *Information Technology and Libraries*, 22(4), 152–157.
- Papavaslopoulos, S., & Poulos, M. (2012). Neural network design and evaluation for classifying library indicators using personal opinion of expert. *Library Management*, 33(4/5), 261–271. <http://doi.org/10.1108/01435121211242308>
- Pu, H., & Yang, C. (2003). Enriching user-oriented class associations for library classification schemes. *The Electronic Library*, 21(2), 130–141. <http://doi.org/10.1108/02640470310470507>
- Robin R. Sewell. (2013). Who is following us? Data mining a library's Twitter followers. *Library Hi Tech*, 31(1), 160–170. <http://doi.org/10.1108/07378831311303994>
- Samson, S. (2014). Usage of e-resources: Virtual value of demographics. *The Journal of Academic Librarianship*, 40(6), 620–625. <http://doi.org/10.1016/j.acalib.2014.10.005>
- Shieh, J.-C. (2012). From website log to findability. *Electronic Library*, The, 30(5), 707–720. <http://doi.org/10.1108/02640471211275747>
- Shreeves, S. L., Kaczmarek, J. S., & Cole, T. W. (2003). Harvesting cultural heritage metadata using the OAI Protocol. *Library Hi Tech*, 21(2), 159–169. <http://doi.org/10.1108/07378830310479802>
- Siriprasoetsin, P., Tuamsuk, K., & Vongprasert, C. (2011). Factors affecting customer relationship management practices in Thai academic libraries. *The International Information & Library Review*, 43(4), 221–229. <http://doi.org/10.1016/j.iilr.2011.10.008>
- Soria, K. M., Fransen, J., & Nackerud, S. (2014). Stacks, Serials, Search Engines, and Students' Success: First-Year Undergraduate Students' Library Use, Academic Achievement, and Retention. *The Journal of Academic Librarianship*, 40(1), 84–91. <http://doi.org/10.1016/j.acalib.2013.12.002>
- Tempelman-Kluit, N., & Pearce, A. (2014). Invoking the User from Data to Design. *College & Research Libraries*, 75(5), 616–640. <http://doi.org/10.5860/crl.75.5.616>
- Todorinova, L., Huse, A., Lewis, B., & Torrence, M. (2011). Making Decisions: Using Electronic Data Collection to Re-Envision Reference Services at the USF Tampa Libraries. *Public Services Quarterly*, 7(1-2), 34–48. <http://doi.org/10.1080/15228959.2011.572780>
- Tosaka, Y., & Weng, C. (2011). Reexamining Content-Enriched Access: Its Effect on Usage and Discovery. *College & Research Libraries*, 72(5), 412–427. <http://doi.org/10.5860/crl-137>
- Walters, W. H. (2007). A Regression-based Approach to Library Fund Allocation. *Library Resources & Technical Services*, 51(4), 263–278.

- Weiner, S. (2009). The Contribution of the Library to the Reputation of a University. *The Journal of Academic Librarianship*, 35(1), 3–13. <http://doi.org/10.1016/j.acalib.2008.10.003>
- Whitmire, E. (2002). Academic library performance measures and undergraduates' library use and educational outcomes. *Library & Information Science Research*, 24(2), 107–128. [http://doi.org/10.1016/S0740-8188\(02\)00108-1](http://doi.org/10.1016/S0740-8188(02)00108-1)
- Will, N. (2006). Data-mining: Improvement of university library services. *Technological Forecasting and Social Change*, 73(8), 1045–1050. <http://doi.org/10.1016/j.techfore.2006.05.006>
- Wu, C.-H. (2003). Data mining applied to material acquisition budget allocation for libraries: design and development. *Expert Systems with Applications*, 25(3), 401–411. [http://doi.org/10.1016/S0957-4174\(03\)00065-4](http://doi.org/10.1016/S0957-4174(03)00065-4)
- Wu, C.-H., Lee, T.-Z., & Kao, S.-C. (2004). Knowledge discovery applied to material acquisitions for libraries. *Information Processing & Management*, 40(4), 709–725. <http://doi.org/10.1016/j.ipm.2003.08.010>
- Yang, S.-T. (2012). An active recommendation approach to improve book-acquisition process. *International Journal of Electronic Business Management*, 10(2), 163–73.
- Yi, Z. (2009). The management of change in information technology: Approaches of academic library directors in the United States. *International Journal of Knowledge, Culture and Change Management*, 9(11), 109–130.
- Yi, Z. (2011). Planning change in the information age: Approaches of academic library directors in the United States. *International Journal of Knowledge, Culture and Change Management*, 10(12), 155–176.
- Yi, Z. (2012). Conducting meetings in the change process: Approaches of academic library directors in the United States. *Library Management*, 33(1/2), 22–35. <http://doi.org/10.1108/01435121211203293>
- Zhang, Q. S., & Wang, X. Y. (2013). Research of Personalized Information Service Based on Association Rules. *Advanced Materials Research*, 760-762, 1800–1803. <http://doi.org/10.4028/www.scientific.net/AMR.760-762.1800>
- Zweibel, S., & Lane, Z. B. (2012). Probing the effects of policy changes by evaluating circulation activity data at Columbia University Libraries. *The Serials Librarian*, 63(1), 17–27. <http://doi.org/10.1080/0361526X.2012.687850>

Curriculum Vitae

Personal details

Name Lorena Catalina Siguenza Guzmán

Birth date and place: January 26, 1980, Cuenca, Ecuador

Nationality Ecuadorian

Education

- Aug. 2015 PhD in Engineering. Department of Mechanical Engineering. Centre for Industrial Management Traffic and Infrastructure. KU Leuven, Belgium. PhD thesis: “Optimal Resource Allocation and Budgeting in Libraries”. Promoter: Prof. Dirk Cattrysse, Co-promoter: Prof. Henri Verhaaren.
- Jun. 2007 Master in Telematics. University of Azuay, Cuenca – Ecuador. MgT thesis: “Desarrollo del prototipo de un sistema de telediagnóstico multimedia bajo el ambiente Web”. Co-author: Laura Ordoñez O., Director: Diego Ponce V. PhD.
- Jun. 2004 Computer Engineer. University of Cuenca, Cuenca – Ecuador. Thesis: “Proceso de informatización del Centro de Información Documental Regional de la Universidad de Cuenca”

Teaching Experience

- Oct. 2010 - ... Aggregate Professor. Department of Computer Science. Faculty of Engineering. University of Cuenca, Ecuador
- 2007 Instructor of the Internetworking Workshop. Telematics Master Program University of Cuenca, Cuenca – Ecuador
- 2007 Lecturer in the module Workshop on Protocols and Networking. Second Edition of the Telematics Master Program. University of Cuenca, Cuenca – Ecuador

2007	Lecturer in the module Workshop on Protocols and Networking. Loja Edition. Telematics Master Program. University of Cuenca, Cuenca – Ecuador
2004 - 2008	Instructor of CCNA1, CCNA2, CCNA3, CCNA4. CISCO Academy. University of Cuenca, Cuenca – Ecuador
Jan 05 – Jul 05	Instructor of the Internetworking Workshop. CEC – EPN, Escuela Politécnica Nacional, Ecuador
2003- 2004	ICT Instructor. University of Cuenca, Cuenca – Ecuador

Professional Experience

Jul. 04 – Oct. 10	IT Engineer. Regional Documentation Centre “Juan Bautista Vazquez”. University of Cuenca, Cuenca – Ecuador
-------------------	--

Research Experience

Oct. 09 – Jul. 15	Research Assistant and PhD Student. Department of Mechanical Engineering. KU Leuven, Belgium
Jul. 14 – Jul. 15	Team Member. Project “Integrating Bibliographic Resources on the Semantic Web”. CEPRA VII, Ecuadorian Consortium for Advanced Internet Development CEDIA
Jan. 08 – ...	Team Member. Project “Institutional change to strengthen research and education”, VLIR – IUC. University of Cuenca, Cuenca – Ecuador

Grants and Recognitions

Dec. 2014	Best Paper Award. Technical Track. 2nd Congreso Ecuatoriano de Tecnologías de Información y Comunicación TIC EC 2015
2013	PhD grant funded by the National Secretary of Higher Education, Science, Technology and Innovation – SENESCYT. Excellence scholarship program. Ecuador.
2009	PhD grant funded by Flemish Inter-University Council (VLIR), Sandwich Program. Belgium.

Miscellaneous

Paper reviews:	Journal of Academic Librarianship (1). Journal of Production Economics (1)
----------------	--

List of publications

Journal Articles

- Siguenza Guzman, L.**, Van den Abbeele, A., Vandewalle, J., Verhaaren, H., Cattrysse, D. (2015). A Holistic Approach to Supporting Academic Libraries in Resource Allocation Processes. *The Library Quarterly: Information, Community, Policy*, 85(3), 295-318. <http://doi.org/10.1086/681612>. Q1 in Scopus.
- Siguenza Guzman, L.**, Saquicela, V., Avila, A., Vandewalle, J., Cattrysse, D. (2015). Literature review of data mining applications in academic libraries. *The Journal of Academic Librarianship*, 41 (4), 499-510. <http://dx.doi.org/10.1016/j.acalib.2015.06.007>. Q1 in Scopus.
- Siguenza Guzman, L.**, Auquilla, A., Van den Abbeele, A., Cattrysse, D. (2015). Using Time-Driven Activity-Based Costing to identify best practices in libraries. *Submitted to the Journal of Academic Librarianship*. Q1 in Scopus.
- Siguenza Guzman, L.**, Van den Abbeele, A., Vandewalle, J., Verhaaren, H., Cattrysse, D. (2014). Using Time-Driven Activity-Based Costing to support Library Management Decisions: A Case Study for Lending and Returning Processes. *The Library Quarterly: Information, Community, Policy*, 84 (1), 76-98. <http://doi.org/10.1086/674032>. Q1 in Scopus.
- Sucozhañay, D., **Siguenza Guzman, L.**, Zhimnay, C., Cattrysse, D., Wyseure, G., De Witte, K., Euwema, M. (2014). Transformational Leadership and Stakeholder Management in Library Change. *The Liber Quarterly*, 24 (2), 55-83. Q3 in Scopus.
- Siguenza Guzman, L.**, Van den Abbeele, A., Cattrysse, D. (2014). Time-Driven Activity-Based Costing Systems for Cataloguing Processes: A Case Study. *The Liber Quarterly*, 23 (3), 160-186. Q3 in Scopus.
- Haro Valle, V., Pérez Rocano, W., **Siguenza Guzman, L.**, Cattrysse, D., Saquicela Galarza, V. (2014). Diseño e Implementación de un Sistema de Soporte de Decisiones para el Centro de Documentación Regional "Juan Bautista Vázquez". *Maskana*, 5 (No. especial - TIC.EC), 245-256.
- Siguenza Guzman, L.**, Van den Abbeele, A., Vandewalle, J., Verhaaren, H., Cattrysse, D. (2013). Recent evolutions in costing systems: A literature review of Time-Driven Activity-Based Costing. *Review of Business and Economic Literature - REBEL*, 58 (1), 34-64.

International Conference Papers

- Siguenza Guzman, L.**, Saquicela, V., Cattrysse, D. (2014). Design of an integrated Decision Support System for library holistic evaluation. Proceedings of IATUL Conferences. 35th Annual IATUL Conference. Helsinki, Finland, 2 - 5 June 2014 (pp. 1-12).
- Siguenza Guzman, L.**, Holans, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., Cattrysse, D. (2013). Towards a holistic analysis tool to support decision-making in libraries. Proceedings of IATUL Conferences. 34th Annual IATUL Conference. Cape Town, South Africa, 14-18 April 2013 (art.nr. 29) (pp. 1-9) Purdue e-Pubs.
- Sucozhañay, D., **Siguenza Guzman, L.**, Zhimnay, C., Cattrysse, D., Euwema, M., De Witte, K. (2011). Change Management strategies for successful implementation of a library 2.0 A case study in Ecuador. INTED2011 Proceedings CD. INTED2011. Valencia, 7-9 March 2011 (pp. 1051-1060). Spain: IATED Publications.

Conference Abstracts

- Siguenza Guzman, L.**, Van den Abbeele, A., Cattrysse, D. (2013). Improving Library Management by Using Cost Analysis Tools: A Case Study for Cataloguing Processes. LIBER Annual Conference. Munich - Germany, 26 - 29 June 2013.
- Sucozhañay, D., **Siguenza Guzman, L.**, Zhimnay, C., Cattrysse, D., Wyseure, G., De Witte, K., Euwema, M. (2011). Managing change in an Ecuadorian library: The role of leaders and stakeholders. WAOP (Werkgemeenschap van onderzoekers in de Arbeids- & Organisatiepsychologie). Tilburg, 25 November 2011.